Human Capital:

Parental Aspirations, Ethnic Differences and Crime

Hieu T.M. Nguyen

A thesis submitted for the degree of Doctor of Philosophy

of The Australian National University

July 2018

Declaration

Unless otherwise indicated this thesis is my own work

Hieu T.M. Nguyen

July 2018

Acknowledgements

This thesis has been completed under the guidance and support from all members of my supervisory panel, to whom I am grateful. Without their assistance, this thesis would not have been possible. I would like to thank my supervisor, Professor Raghbendra Jha, for his invaluable guidance and constant support. I deeply appreciate all the sympathy and encouragement that he has given me since the first days of my PhD program. Special thanks to Associate Professor Blane D Lewis, my supervisor, who always enthusiastically shows me other ways to go whenever my research does not progress as expected. His insightful suggestions and direction, which I am lucky enough to receive throughout this research work, have greatly contributed to the completion of my thesis. I also thank Creina Day, my associate supervisor.

In writing the thesis, I have benefited from discussions with Donny Pasaribu, Giang Nguyen and Moh Agung Widodo. Valuable comments from lecturers in Arndt-Corden Department of Economics and participants in the Economics PhD Seminars are greatly appreciated. I also acknowledge constructive comments from the two anonymous reviewers of Journal Education Economics, in which I published one of my papers. Advice on academic writing from Dr Megan Poore has greatly improved the presentation of this thesis.

I gratefully acknowledge the financial support from Endeavour Scholarship, which administered by the Australian Government Department of Education and Training, to undertake my PhD. I also acknowledge HDR Student Conference Funding - The Crawford School of Public Policy and The Vice-Chancellor's HDR Travel Grants,

both of which funded my presentation of the 5th chapter of this thesis at the XXVII Meeting of the Economics of Education Association in Spain, June 2018.

Finally, I take this opportunity to express my deepest gratitude to my beloved mother and mother-in-law for their unconditional love, care, and sacrifice. Without them, I would not be where I am today. I extend my profound gratitude to my father who always looked forward to the day I could graduate, fostering my motivation. I warmly thank my sister Thuy, brother in-law Dung, brother Trung and in-law sister Lien for their constant support and, especially, taking care of my parents, so that I can pursue my PhD program. Heartfelt thanks belong to my dearest husband and son, who are always with me to nurture and pursue our dreams. I dedicate this work to you, my family!

Abstract

This thesis comprises three empirical papers. Using various econometric techniques, each paper examines different aspects of human capital. These aspects vary from determinants of human capital to heterogeneity in human capital, and to non-economic benefit of human capital. The objective of this thesis is to obtain a better understanding about issues related human capital and education in developing countries, in particular Indonesia and Vietnam.

The first research paper examines the role of parental academic aspirations for a child in determining the allocation of the child's time. Parental academic aspirations are included in a theoretical model of household utility maximization to assess how they affect the child's optimal time allocation. Then, an empirical analysis is conducted to verify the effect of parental aspirations on a child's hours of study and work by using data from the 2006 and 2014 Vietnam Young Lives Survey. The instrumental variable approach is adopted to address a problem related to the simultaneity of decision making process and endogeneity of parental aspirations. Both the theoretical and empirical results show that parental academic aspirations are positively associated with the child's hours of study and negatively associated with the child's hours of work.

The second research paper seeks to understand the reasons for the ethnic gaps in education outcomes, measured by enrolment rate, schooling progress and test scores, in Vietnam. The examination employs Probit and multilevel regression models, and associated decomposition techniques. The estimation results suggest that the factors mainly driving the poor performance at school of minority children are not their own internal attributes, but disadvantages related to the external

determinants. The decomposition results show that the gap in enrolment is mostly explained by household characteristics. However, all child, household and commune attributes significantly contribute to the gap in schooling progress. The test score gaps are attributable to a broader set of variables such as parent's education, the use of the Vietnamese language, peer and school characteristics.

The third paper identifies the causal effect of education on property crime at the district level in Indonesia over the period 2007–2012. Both dynamics of crime and endogeneity of education are taken into accounted by applying difference generalized method of moments. The results show that more educated neighbourhoods experience less crime. Secondary and higher education play a particularly important role in crime reduction. Effects are more pronounced for crimes reported by males than females. One mechanism seems to be that there are fewer opportunities for engaging in criminal behaviour when one is in school. Extreme poverty appears to weaken the extent to which education reduces crime.

Tables of contents

List of tables

List of figures

List of abbreviations

ADB	Asian Development Bank
BMI	Business Monitor International
BMI	Body Mass Index
BPS	Statistics Indonesia
GDP	Gross Domestic Production
GMM	Generalised Method of Moments
GSO	Vietnam General Statistics Office
INDO-DAPOER	The Indonesia Database for Policy and Economic Research
IQ	Intelligence Quotient
IRT	Item Response Theory
MOEC	Ministry of Education and Culture
MOET	Ministry of Education and Training
MPC	Minnesota Population Center
OECD	Organisation for Economic Co-operation and Development
PPVT	Peabody Picture Vocabulary Test
SAGE	Schooling for age index
Susenas	Indonesia Social and Economic Surveys
TI	Transparency International
UIS	The United Nations Educational, Scientific and Cultural Organisation Institute for Statistics
UNESCO	The United Nations Educational, Scientific and Cultural Organisation

UNICEF	United Nations International Children's Emergency Fund
VHLSS	Vietnam Household Living Standard Survey
VND	Vietnamese currency
WB	The World Bank
WHO	The World Health Organization
WISC	Wechsler Intelligence Scale for Children
YL	Young Lives
YLS	Vietnam Young Lives Survey
YLSS	Vietnam Young Lives School Survey

1. Introduction

Human capital has been widely considered as a factor fuelling economic growth through increasing labour productivity. In many studies, human capital has been even recognized as an important determinant of the economic development level (Hanushek and Woessmann 2008; Gennaioli et al. 2012; Cinnirella and Streb 2017). Besides, non-economic benefits of human capital have also recently been acknowledged. Human capital helps to improve health outcomes (Silles 2009; Oreopoulos 2006; de Walque 2010; Grimard and Parent 2007), increases political and social participation (Milligan, Moretti, and Oreopoulos 2004; Dee 2004) and reduces risky behaviours, such as crime and teenage pregnancy (Bell, Costa, and Machin 2016; Berthelon and Kruger 2011; Lochner and Moretti 2004; Machin, Marie, and Vujić 2012; Cook and Kang 2016; Black, Devereux, and Salvanes 2011).

Human capital is broadly defined as individuals' knowledge and skills, which are accumulated through education, training and working experience, as well as the innate ability. However, in most of the studies, human capital in a narrow definition is measured by education levels of people (Gennaioli et al. 2012; Acemoglu, Gallego, and Robinson 2014; Benhabib and Spiegel 1994). In either a broad or narrow definition, human capital still concerns a wide range of issues, which cannot be thoroughly covered in a thesis. Therefore, this thesis focuses only on three specific issues related to human capital: the allocation of a child's time, inequality in education, and the crime reducing effect of education. The aim of this thesis is to gain a better understanding about these three aspects of human capital in general,

and that in Indonesia and Vietnam in particular. Based on this understanding, the thesis proposes policy implications, which are relevant to the context of each country studied, to improve education outcomes and to facilitate the development of human capital.

The thesis consists of six chapters. Chapter 1 provides a general introduction about the thesis. Chapter 2 details background on education in Indonesia and Vietnam. Using data from Vietnam, Chapter 3 examines the allocation of a child's time between study and work by focusing on the role of parental academic aspirations. Chapter 4 investigates ethnic gaps in education outcomes among the young generation in Vietnam. Chapter 5 explores a non-economic benefit of education, crime reducing effect, in Indonesia. Chapter 6 concludes.

1.1. Parental academic aspirations for a child and the allocation of a child's time

Since human capital is accumulated gradually through learning and working, the way children spend their time on studying, working and leisure directly affects their physical and mental health, their skills and knowledge, thus determines their human capital. In Chapter 3, the allocation of a child's time between study and work is examined by emphasizing the role of parental academic aspirations for their child. Although a child's time allocation has been investigated in association with a numbers of factors in the literature, most of them are socio-economic and demographic attributes. The connection between a child's time allocation and psychological factors, such as aspirations and expectations, has not been seriously attended to, thus motivating the analysis conducted in Chapter 3.

In Chapter 3, parental academic aspirations are introduced in a theoretical model of household utility maximization to feature how they affect the optimal child's time allocation. In this model, parental aspirations are considered as a factor augmenting the effectiveness of the child's study time, which, in turn, increases the child's human capital. Then an empirical analysis is conducted to verify the effect of parental aspirations on a child's hours of study and work by using data from the 2006 and 2014 Vietnam Young Lives Survey. To address the problems related to the simultaneity of decision making process and endogeneity of parental aspirations, the average local wage is used as an instrumental variable for aspirations. Both the theoretical and empirical results show that children, whose parents have high academic aspirations for, spend more time studying and less time doing economic work. A further analysis suggests that parental academic aspirations have no significant effect on the time doing housework for boys, but significantly reduce the workload for girls. Academic performance of a child is negatively associated with the number hours of work and positively associated with the time he or she spends studying.

The consistent findings of the role of parental aspirations in determining a child's studying and working time from both theoretical and empirical analyses suggest that the allocation of the child's time is a missing link in the relationship between parental aspirations and the child's academic achievements. Thus, future studies might consider the allocation of a child's time as a channel among others, through which parental aspirations affect a child's educational attainment.

1.2. Ethnic gaps in education outcomes in Vietnam

Chapter 4 focuses on the narrow definition of human capital, which is education. More specifically, the second research paper presented in Chapter 4 attempts to explain why education outcomes vary between children in non-Kinh minority and Kinh majority groups in Vietnam. Although the goal of a fair society has been set in Vietnam since independence in 1945, inequality including that in education is an obstacle to the country achieving its goal. For example, ethnic gaps in education exist in all education outcomes. Non-Kinh minority children lag far behind Kinh majority counterparts in enrolment rates, schooling progress and test scores. Therefore, identifying the key factors driving the gaps will not only help to narrow the gaps, but also to improve the education levels of people in the country and contribute to the development of a fairer society. That is also the aim of the research paper presented in Chapter 4.

In particular, the paper explains the ethnic gaps in school enrolment, schooling progress and academic performance by using data from the 2009 Young Lives Survey and 2011-2012 Young Lives School Survey. The analysis employs Probit and multilevel regression models, and associated decomposition techniques. The estimation results suggest that the factors mainly driving the poor performance at school of minority children are not their own internal attributes, but disadvantages related to the external determinants. Poverty is still an obstacle to minority children obtaining high test scores. Education of the father, who is likely to be the decision maker in the family, significantly affects the enrolment status of the child, while education of the mother, who tends to spend more time with the child, is substantially important to the child schooling progress. The decomposition

results show that the key factors contributing to the gaps are different, depending on the type of education outcomes under examination. The gap in enrolment is mostly explained by household characteristics, of which father's education and household economic conditions are particularly important. However, all child, household and commune attributes significantly contribute to the gap in schooling progress. The test score gaps are attributable to a broader set of variables such as parent's education, the use of the Vietnamese language, peer and school characteristics.

The findings related to the variation in the role of determinants suggest that efforts to narrow the ethnic gaps in education should vary the focus and priority according to the targeted outcomes. For example, improving household economic condition might be relevant to narrowing the enrolment and schooling progress gaps, but in order to equalize education performance across ethnic groups, removing the language barriers and improving school quality are crucial.

1.3. The crime reducing impact of education in Indonesia

Chapter 5 looks at a non-economic benefit of human capital: crime reducing effect. In this chapter, human capital is measured by education levels of people. Evidence of the crime reducing effect of education has been found in a number of developed countries, but the topic is studied in developing countries only in a limited manner, where law enforcement tends to be weaker, poverty higher and the average education level of the population lower. Indonesia is an interesting case because crime rates trend to decrease recently, despite limited capacity and insufficiency of the police force, and serious corruption in the justice system. Given a consistent

improvement in education levels of Indonesians over time, it is reasonable to ask whether education plays a role in the reduction of crime incidence in the country.

More precisely, the chapter identifies the causal effect of education on property crime at the district level in Indonesia over the period 2007–2012. It establishes how the education level of people living in a district affects the incidence of property crime they have experienced. Hence, information on crime comes from self-reports of individuals who were victims of crime rather than from police or offender sources, which are likely to undercount the incidence of crime. Both dynamics of crime and endogeneity of education are taken into accounted by applying difference generalized method of moments. In addition, the paper examines the effect of education on crime in a developing country context This has been inadequately studied so far. The results show that more educated neighbourhoods experience less crime. Secondary and higher education play a particularly important role in crime reduction. Effects are more pronounced for crimes reported by males than females. One mechanism seems to be that there are fewer opportunities for engaging in criminal behaviour when one is in school. Extreme poverty appears to weaken the extent to which education reduces crime.

The findings relating to the crime reducing impact of education (especially senior secondary education) in this paper provide an additional supporting reason for the implementation of compulsory senior secondary education in Indonesia, which was commenced recently. A key recommendation of this paper is that the non-pecuniary benefits of schooling, such as the crime reducing effect, should be taken into account in the policy making process, for example policies related to the expansion of compulsory education through senior secondary level. Additionally,

the government should consider educational improvement and poverty mitigation as possible solutions, apart from law enforcement, to crime.

2. Background

2.1. Education in Indonesia

With the population of over 260 million, Indonesia has the 4th largest education system in the world, following China, India, and the United States. At the pre-tertiary level, there are 250 thousand schools, 50 million students, and 2.6 million teachers nationwide. The ratio of students to teachers in primary education is 18.6 (OECD and ADB 2015). Basic education consists of primary, junior secondary and senior secondary level. Primary education in Indonesia includes 6 years of schooling, junior and senior secondary each spans 3 years (Cerdan-Infantes et al. 2013). Public and private schools dominate in the education system, accounting for 84 per cent of the total. Private schools play an important role, especially in secondary education. Specifically, the share of private schools is 7 per cent at primary, 56 per cent at junior secondary and 67 per cent at senior secondary level. Islamic schools account for the remaining 16 per cent of basis education system. Ministry of Education and Culture has overall responsibility for managing public and private schools while the Ministry of Religion oversees Islamic schools (OECD and ADB 2015).

Before 1979, the academic and calendar years were coincident. In 1979 the school year began to operate from July and run through June the following year. Students who started the school year in January of 1978 remained in the same grade until July of 1979 (Parinduri 2014). Government regulations deem that children

should start primary school at age seven, although some begin a year earlier or a year later (Barakat 2016).

Prior to 2001 schools were managed by central government ministries and their geographically deconcentrated field offices. In 2001 Indonesia decentralized authority over many public services of a local nature, including education, to the local government at provincial and district level. At that time districts became responsible for financing and managing early childhood, primary, junior secondary, and senior secondary public schools (Lewis 2014). In 2009, for example, district government spending on education accounted for more than half of total spending on education by all levels of government, while the proportion of that spending by provincial government was just around 5 per cent. In 2016, financial and managerial authority for public senior secondary schools was reassigned to the provincial level. Islamic schools continue to be administered by the Ministry of Religion.

In 2003, Education Law 20/2003 was promulgated to specify that 20 per cent of national and subnational budgets should be devoted to expenditure on education. It effectively mandates all provinces and districts to spend a minimum of 20 per cent of their budgets on education (OECD and ADB 2015). The expenditure mandate has led to large increases in public education spending and student enrolments in junior and senior secondary school have risen significantly as a result (Cerdan-Infantes et al. 2013).

In 1984 the government implemented its National Compulsory Education program, requiring all children to complete primary school. The six-year compulsory education initiative, along with the school construction program implemented during 1970s, played a strong role in increasing enrolments at the

primary level and improving educational attainment more generally (Suryadarma et al. 2006). By the beginning of 1990s primary education was almost universal across the country, and the net enrolment rate at primary education in Indonesia reached over 92 per cent in 1994 (BPS 2018). Based in part on this perceived success, the government introduced 9 year compulsory schooling in 1994 to increase access to junior secondary education. After 20 years of implementing 9 year basic education, the net enrolment rates at junior secondary level increased by almost 30 percentage points, from 50 per cent in 1994 to 77 per cent in 2014. In 2015, Indonesia extended compulsory schooling to include senior secondary school, for which the net enrolment rate was still low, at 59 per cent.

All the above mentioned efforts made by Indonesian government have significantly contributed to the improvement in the education levels of people in the country. The share of the population aged 15 and above never attending school decreased to 3 per cent in 2017, from 14 per cent in 1994. The percentage of population without education qualification reduced by half in the same period, from 22 to 11 per cent (BPS 2018). There is almost gender equality in the enrolment rate in Indonesia, with the gender parity index of female to male students was 0.98 in 2009/10 (OECD and ADB 2015).

Despite these achievements, the educational gaps between provinces, rural-urban areas, and rich and poor families still exist in Indonesia. The gaps are especially considerable at junior secondary level and higher. For example, in 2014 the net enrolment rate at junior secondary school was 95 per cent in Jakarta, compared to 32 per cent in Papua (MOEC 2013). Similarly, the difference in the rate between urban and rural is more than 10 percentage points (UNICEF cited in

OECD and ADB 2015). With regard to economic condition, the dropout rate of children from the poorest quintile households is almost 5 times more than that of those from the riches households (UNICEF cited in OECD and ADB 2015).

2.2. Education in Vietnam

Like Indonesia, the basic education in Vietnam comprises 12 years of schooling: 5 years of primary level, 4 years of lower secondary and 3 years of upper secondary level. In the school year 2014-2015, there were over 861 thousand teachers and 15.3 million students in the country. The number of students per teacher is 19.6 for primary education, 16.3 for lower secondary and 15.9 for upper secondary education. The education system consists of nearly 29 thousand schools, including over 15 thousand primary, 10 thousand lower secondary, almost 2.5 thousand upper secondary, and the remainder comprise schools with combined levels of education (GSO 2016). There is at least one primary school located in every commune and almost every commune has a lower secondary school. All districts have an upper secondary school (MOET 2015).

In Vietnam, the school year starts on 5th September and lasts until the end of May in the following year. In the Education Law, which was first promulgated in 1998 and then amended in 2005 and 2009, the school starting age for children in Vietnam is six year old. In particular, children enter grade 1 in September of the calendar year in which they turn six. If students enter school as the regulation age and continue their school progress properly without any interruption or grade repetition, they will finish grade 12 in the year in which they turn 18.

The Ministry of Education and Training (MOET) has overall responsibility for managing education in Vietnam. In general, administration of the education system is quite centralized (Jonathan 2011). However, decentralization, which has been implemented strongly since 2000s, provides education institutions with significant decision-making authority. Decree No. 10/2002/ND-CP and Decree No. 43/2006/ND-CP are two examples of government's documents that enhance autonomy of public schools and other non-profit service providers.

Vietnamese government expenditure on education has become among the highest in Asia (MOET, GSO, and UIS 2016). Government spending on education was just one per cent of GDP in 1990 (Jonathan 2011) but rose to 4.9 per cent of GDP in 2008 and 5.7 per cent of GDP in 2013[1]. In 2013, government education expenditure accounted for 20 per cent of total spending, increasing from 16 per cent in 2009. Expenditure on primary education makes up the largest share of the total government and household expenditure on education, around 25 per cent, followed by that on secondary and university levels, which individually accounts for roughly 20 per cent of the total education expenditure (MOET, GSO, and UIS 2016). Like other developing countries most public spending on education, about 80 per cent of recurrent budgets, is allocated to teacher salaries (Jonathan 2011).

Vietnam achieved universal primary education in 2000. By 2013, the net primary school enrolment rate had risen to 97 per cent (MOET 2015). After achieving universal primary education, the government issued degree 88/2001/NĐ-CP in 2001 to set the target of a universal access to lower secondary education for all children aged 11-18 by 2010. In the school year 2008-2009, however, all

[1] https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?locations=VN

districts and provinces in the country reported that they had already reached the objective (MOET 2015). The net enrolment rates for lower secondary education, therefore, have been rapidly increased, from 30 per cent in 1993 to 79 per cent in 2008 (Jonathan 2011) and 88 per cent in 2013 (MOET 2015). The net enrolment rates for upper secondary education surged to 54 per cent in 2008 from 7 per cent in 1993 (Jonathan 2011). Some provinces already initiated programs to reach universal upper secondary education (MOET 2015).

Although Vietnam has made remarkable progress in improving average access to schooling, there are still considerable inequalities in educational attainment, especially between children living in urban and rural areas and between those belonging to Kinh and non-Kinh ethnic groups. The net enrolment rates for urban and rural areas in 2010, for example, were 86 and 79.7 per cent for lower secondary, and 69.6 and 55.4 per cent for upper secondary, respectively (GSO 2011). Differences in enrolment rates at primary, lower secondary, higher secondary schools and university between Kinh and non-Kinh children were 8, 26, 35 and 18 per cent, respectively (GSO 2010a). Similarly, educational gaps between children in low income and high-income family exist across age groups. For example, more than a half of children who belong to the 1st income quintiles drop out school when they are 15-17 year old, compared to 15.6 per cent of those in the 5th quintile (Quyen 2011).

On a more positive note, Vietnam is moving toward gender equality in education. The gender gap in favour of boys has been gradually removed, and in some education indicators, girls even outperform boys. In particular, the net enrolment rate in secondary schools for girls was 5 per cent lower than boys in

1990. However, by 2010, enrolment rates for girls exceeded that for boys. For instance, the enrolment rates were 82.6 per cent and 80.1 per cent for lower secondary education, and 63.1 per cent and 53.7 per cent and for upper secondary education in 2010, respectively (Jonathan 2011).

3. Effect of parental academic aspirations for a child on the child's time allocation

3.1. Introduction

Determinants of children's education outcomes have been widely studied in both educational economics and psychology. Economists acknowledge that a child's time allocation is an important determinant of educational attainment. For example, a positive association between study hours and learning outcomes has been documented (Hacker et al. 2000; Stanca 2006; Chan, Shum, and Wright 1997), although a causal relationship between them remains unclear. The evidence of adverse effects that child labour exerts on schooling, in contrast, is uncontroversial (Gunnarsson, Orazem, and Sánchez 2006; Heady 2003; Lillydahl 1990; Psacharopoulos 1997; Ray and Lancaster 2005; Zabaleta 2011).

In the realm of psychology, parental academic aspirations for a child are found to be a robust factor affecting the child's academic achievement (Areepattamannil and Lee 2014; Creed, Conlon, and Zimmer-Gembeck 2007; Schoon, Parsons, and Sacker 2004; Strand 2011; Zhang et al. 2011). Psychology research has identified the channel of this effect as the association between parental aspirations, and emotional and literacy support that parents provide to their child, which, in turn, facilitates the child's educational attainment. For example, Davis-Kean (2005) constructed a model in which academic expectations of parents for their offspring determined their behavior such as reading and playing with their child, and thus influenced the child's cognitive scores. Likewise, Froiland,

Peterson, and Davison (2013) highlighted that parental aspirations indirectly affected the learning outcomes of a child through influencing the child's aspirations and parents' supportive behavior, e.g. shared reading with the child. A similar conclusion was made by Christofides et al. (2015) and Favara (2017) when they both found transmission from parents' aspirations to their children.

The indirect impact of parental aspirations and the direct effect of a child's time allocation on the child's academic achievement found in the existing literature together raise a question about whether a child's time allocation is a missing link in the relationship between parental aspirations and the child's academic achievement. Is it possible that parental academic aspirations for a child determine the way that parents allocate their child's time, which, in turn, affects the child's educational attainment? If this is the case, time allocation can be considered as a channel through which parental aspirations affect education outcomes of a child.

Since the impacts of parental aspirations and time allocation on children's education outcomes have received significant attention, they are not the main focus of this paper. Instead, this paper investigates the role of parental academic aspirations in determining a child's time allocation. To do this, the traditional model of household utility maximization is extended by introducing parental aspirations into the child's human capital function. Next, data from the 2006 and 2014 Vietnam Young Lives Survey are employed to test the hypothesis about a positive association between parental aspirations and a child's hours of study, and a negative relationship between parental aspirations and a child's work hours. This paper extends the literature by theoretically and empirically investigating how parental academic aspirations for a child affect the child's time allocation.

3.2. Determinants of a child's time allocation in the literature

In the literature, a child's time allocation has been examined in the association with various socio-economic factors. However, none of the studies up to date have looked at the relationship between parental academic aspirations for a child and the allocation of a child's time.

Poverty is perhaps the most acknowledged determinant of a child's work and study time (Bacolod and Ranjan 2008; Edmonds 2005; Edmonds and Turk 2002; Shafiq 2007; Soares, Kruger, and Berthelon 2012; Edmonds 2006a; Malik 2013; Ray 2000, 2002). The overall findings demonstrate that poverty increases the former and decreases the latter. Nevertheless, there are some exceptions to this general pattern, for instance, independence of school enrollment rate and wealth (Hou 2010), and a positive association between child labor and agricultural land area (Bhalotra and Heady 2003; Lima, Mesquita, and Wanamaker 2015).

Apart from the household's economic condition, both parents' education and a child's gender are robust predictors of a child's activities. Specifically, parents' education significantly reduces both school dropouts and child labor (Chevalier et al. 2013; Kumar 2015; Fan and Chen 2001; Mukherjee and Das 2008; Shafiq 2007). Concerning gender difference, girls have a higher probability of participating in domestic work and a lower probability of being involved in paid market work than boys (Ilahi 2001; Edmonds 2008); girls are more vulnerable to a household's welfare shocks than boys are (Ilahi 2001); the background of a parent of the same gender as the child has stronger effects on the child's study and work (Emerson and Souza 2002).

With regards to demographic factors, birth order of a child and number of siblings are found to significantly determine how the child spend her or his time (Dammert 2010; Dang and Rogers 2013; Edmonds 2006b; Ejrnæs and Pörtner 2004; Emerson and Souza 2008; Hong 2013; Seid 2013). While results conclusively show that first-born children are more likely to work than later-born children, the impacts of birth order on school enrollment are somewhat puzzling. In addition, a child's time allocation is also considered as an outcome of intra-household bargaining (Kambhampati 2009; Ridao-Cano 2001). However, the findings are mixed. Kambhampati (2009) found that a mother's bargaining power, measured by her contribution to household expenditures, appeared to be negatively associated with both schooling and child labor. Ridao-Cano (2001) used a mother's education and access to credit to the represent mother's bargaining power in a family. The author concluded that there was a positive association between the relative bargaining power of a mother to father and a rise studying time of the child.

Other determinants that have been examined in the literature include remittance and credit (Alcaraz, Chiquiar, and Salcedo 2012; Bouoiyour and Miftah 2014; Islam and Choe 2013), trade liberalization and globalization (Chaudhuri 2004; Kis-Katos and Sparrow 2011), geographic differences including regional and urban-rural differences (Chamarbagwala 2008; Ersado 2002; Liu 2000), country differences (Gunnarsson, Orazem, and Sánchez 2006; Maitra and Ray 2002; Ray and Lancaster 2005; Rosati and Rossi 2001), and market factors (Chakrabarty, Grote, and Lüchters 2011; Du 2013; Ersado 2005; Hong 2013; Kruger 2007; Skoufias 1993, 1994). Nonetheless, no research has yet explored the effects of

parental academic aspirations for the child on a child's time allocation. This study, to the best of the author's knowledge, is the first to do so.

3.3. Theoretical model

In economics, the allocation of a child's time has been considered as the choice of parents in the context of maximizing household's utility (Baland and Robinson 2000; Basu and Van 1998; Cigno and Rosati 2005; Skoufias 1993). In order to examine how parental academic aspirations for a child affect the child's time allocation, this section extends the model of a household's behaviour by introducing parental aspirations into the human capital function of the child.

Assume that a household consists of a parent and a child, and the child has 1 unit of time to spend on studying and working. For simplicity, leisure time is considered to be constant, and market work and home production are not separated. In this household, the parent makes all decisions on allocating the child's time to maximize the household's utility. The common utility function of the household, as shown in equation (3.1), is a function of consumed goods, *c*, and human capital of the child, *h*. The household's utility function is twice-differentiable and quasi-concave in consumption, and a linear function of human capital. The human capital that the child obtains, in turn, is a function of study time of the child, t_s. The human capital function has features of a production function, which is twice differentiable and quasi-concave in its argument, t_s. In the human capital function, parental academic aspirations for the child, *A*, help to augment the effectiveness of the child's study time. The unit cost of the child's study time is *s*, e.g. school fees, and

the wage rate per unit of the child's work time is *w*. For the sake of simplicity, the parent's income is assumed to be exogenous in the model and is denoted as *y*. The income of the child and parent together generate the full income of the household, which, then, is used to expend on goods, *c*. The price of consumed commodities is normalized to 1.

The specified model is as follows.

The household's utility function:

$$u = \ln c + h \Rightarrow \max \tag{3.1}$$

The utility function *u* in (3.1), as mentioned above, is twice-differentiable and quasi-concave in consumption *c*, and a linear function of human capital *h*. Equation (3.1) satisfies the conditions: $u'(c) > 0$, $u''(c) < 0$, $u'(h) = \text{constant}$ and $u'(h) > 0$.

The human capital function of the child:

$$h = (At_s)^{\alpha} \tag{3.2}$$

where, $A > 0$, $0 < \alpha < 1$. Equation (3.2) satisfies the conditions: $h'(t_s) > 0$, $h''(t_s) < 0$.

The household's budget constraint[2]:

$$c = y + wt_w - st_s \tag{3.3}$$

2. If the hours of housework are taken into account, the budget constraint can be written: $c = y + (t_0 + \delta t_h)w_p + wt_w - st_s$, where y is the non-wage income of the household, t_h is the child's hours of housework, t_0 is the minimum hours that the parent works for wage, w_p is the unit wage of the parent and $0 < \delta < 1$. Because the child spends t_h hours to help the parent with housework, the parent can spend more time, δt_h, on paid work. In this case, it is possible to examine the combination of the child's economic work and housework hours.

The child's time constraint:

$$t_s + t_w = 1 \tag{3.4}$$

The non-negativity constraints:

$$t_w \geq 0\,,\ t_s \geq 0 \tag{3.5}$$

Substitution of (3.2), (3.3), (3.4) into (3.1) yields the following model:

$$U = \ln(y + wt_w - st_s) + (At_s)^{\alpha} \Rightarrow \max \tag{3.6}$$

Subject to:

$$t_s + t_w = 1\,, t_w \geq 0\,,\ t_s \geq 0 \tag{3.7}$$

To solve the optimization problem shown in equations (3.6) and (3.7), the Lagrangian function is formed:

$$L = \ln(y + wt_w - st_s) + (At_s)^{\alpha} + \lambda(1 - t_w - t_s) + \mu_w t_w + \mu_s t_s \tag{3.8}$$

Differentiating the Lagrangian function with respect to t_s, t_w, we obtain the Kuhn-Tucker conditions:

$$\frac{\partial L}{\partial t_w} = U'(t_w) - \lambda + \mu_w = \frac{w}{y + wt_w - st_s} - \lambda + \mu_w = 0 \tag{3.9}$$

$$\frac{\partial L}{\partial t_s} = U'(t_s) - \lambda + \mu_s = -\frac{s}{y + wt_w - st_s} + \alpha A^{\alpha}(t_s)^{\alpha-1} - \lambda + \mu_s = 0 \tag{3.10}$$

$$\frac{\partial L}{\partial \lambda} = 1 - t_w - t_s = 0 \tag{3.11}$$

The complementary slackness conditions are:

$$\mu_w t_w = 0\,;\ \mu_w \geq 0\,;\ t_w \geq 0 \tag{3.12}$$

$$\mu_s t_s = 0\,;\ \mu_s \geq 0\,;\ t_s \geq 0 \tag{3.13}$$

The Kuhn-Tucker conditions help to explain why a child specializes in an activity and does not take part in another activity.

For example, if the child participates in both activities, $t_w > 0$ and $t_s > 0$ (the interior solution), it can be derived from (3.12) and (3.13) that $\mu_w = \mu_s = 0$. Thus, (3.9) and (3.10) can be rewritten as follows.

$$\lambda = U'(t_w) = U'(t_s) \tag{3.14}$$

Equation (3.14) implies that the child is engaged in both activities if the marginal return to work equals the marginal return to study.

Likewise, in the case that the child specializes in study, $t_w = 0$ and $t_s > 0$, then $\mu_w > 0$ and $\mu_s = 0$. Thus, $\lambda = U'(t_w) + \mu_w = U'(t_s)$; alternatively, this can be rewritten as follows.

$$U'(t_w) \le U'(t_s) \tag{3.15}$$

Expression (3.15) shows that the child specializes in study if the marginal return to study at least equals the marginal return to work.

Similarly, an explanation for the case when the child specializes in work can be derived.

Because *w*, *y* and *A* are exogenous variables, solving the Kuhn-Tucker conditions provides the solution to the child's optimal time allocation, which can be written in reduced form equations:

$$t_s = t_s(w, y, A) \tag{3.16}$$

$$t_w = t_w(w, y, A) \tag{3.17}$$

To examine the impact of parental academic aspirations for the child on the child's time allocation, let's consider the inner solution, where $t_w > 0$, $t_s > 0$ and $\mu_w = \mu_s = 0$. It can be derived from (3.9) and (3.10) that

$$\alpha A^{\alpha} (t_s)^{\alpha-1} \left[y + w t_w - s t_s \right] - w - s = 0 \tag{3.18}$$

Substitution of (3.11) into (3.18) results in (3.19)

$$\alpha A^{\alpha}(t_s)^{\alpha-1}\left[y+w-t_s(w+s)\right]-w-s=0 \tag{3.19}$$

The impact of parental academic aspirations can be evaluated without any requirement of the explicit solution to work hours and study hours. To do so, the next step is to find the total derivative of the equation (3.19).

$$\left\{\alpha^2(At_s)^{\alpha-1}\left[y+w-t_s(w+s)\right]\right\}dA+\left\{-\alpha A^{\alpha}(t_s)^{\alpha-1}(w+s)+\alpha(\alpha-1)A^{\alpha}(t_s)^{\alpha-2}\left[y+w-t_s(w+s)\right]\right\}dt_s=0 \tag{3.20}$$

Transforming equation (3.20) to obtain dt_s/dA in one side of the equation as follows.

$$\frac{dt_s}{dA}=\frac{-\alpha^2(At_s)^{\alpha-1}\left[y+w-t_s(w+s)\right]}{-\alpha A^{\alpha}(t_s)^{\alpha-1}(w+s)+\alpha(\alpha-1)A^{\alpha}(t_s)^{\alpha-2}\left[y+w-t_s(w+s)\right]}>0 \tag{3.21}$$

In equation (3.21), $dt_s/dA>0$ due to $y>st_s$, $w.1>wt_s$ and $0<\alpha<1$

$$\frac{dt_w}{dA}=\frac{d(1-t_s)}{dA}=-\frac{dt_s}{dA}<0 \tag{3.22}$$

Expressions (3.21) and (3.22) show that a rise in parental academic aspirations for a child results in an increase in study time and a decrease in work time of the child.

3.4. Data

The Vietnam Young Lives Survey (YLS) tracks data on 2000 children who were born in 2001/2002 (younger cohort) and 1000 children who were born in 1994/1995 (older cohort) over a 15-year period. Four rounds of the survey were implemented in 2002, 2006, 2009 and 2014. The data were collected in 5 provinces representing

5 out of 9 economic regions[3]. The selected provinces were Lao Cai in the North East region, Hung Yen in the Red River Delta, Da Nang in the City region, Phu Yen in the South Central Coast region and Ben Tre in the Mekong River Delta region.

Although the YLS contains rich information on children's well-being, the question of parental academic aspirations for their child was only asked in round 1 (2002) and round 2 (2006) for the older cohort, and in round 2, round 3 (2009) and round 4 (2014) for the younger cohort. Information on the time that a child spent studying, working, and on leisure was only available in round 2 and afterward. Of children in the younger cohort, none worked in 2006, when they were 5 years old, and just 5 per cent of them worked in 2009, when they were 8 years old. Given the unavailability of information relating to parental academic aspirations for their children and the lack of variation in hours of work, the empirical analysis in this paper includes only children from the old cohort in round 2 and the young cohort in round 4. After dropping observations with missing data, the final sample consists of 913 children born in 1994-1995, and 1812 children born in 2000-2011. The age of children in both cohorts ranges from 11 to 13 year old.

3. In the Young Lives project, Vietnam is divided into 9 regions: North-West, North-East, Red River Delta, North Central Coast, South Central Coast, South-East, Central Highlands, Mekong River Delta and City region, which includes the 5 major cities of Vietnam (Hanoi, Ho Chi Minh City, Da Nang, Hai Phong, and Ba Ria-Vung Tau)

3.5. Empirical strategy

In the theoretical model, the parent makes a decision on allocating the child's time between studying and working in order to maximize the household's utility. The solution to this problem can be presented by a system of reduced form equations, in which study and work time are left-hand side variables, and parental academic aspirations for the child appear on the right-hand side of the equations. Empirical strategy is, therefore, related to identifying a proper approach to estimate this system of reduced form equations.

3.5.1. Variable selection

Dependent variables in the equations of time allocation are the number of daily hours that a child spends doing economic work and studying. Hours of economic work include the time that a child works for pay and works at household business without payment. From this definition, the time that a child takes care of younger siblings or does chores is not included in hours of economic work. The time that a child studies at home and at school together constitutes studying time. Figure 3.1 and Figure 3.2 respectively present histograms of hours of study and economic work by gender and cohort. In general, girls and the younger cohort have more hours of study and fewer hours of economic work than their counterpart children. Most children spend 6 to 9 hours per day studying, and approximately 3 per cent of children do not study at all. The proportion of economically active children is around 22 per cent and the working time is usually from 1 to 4 hours.

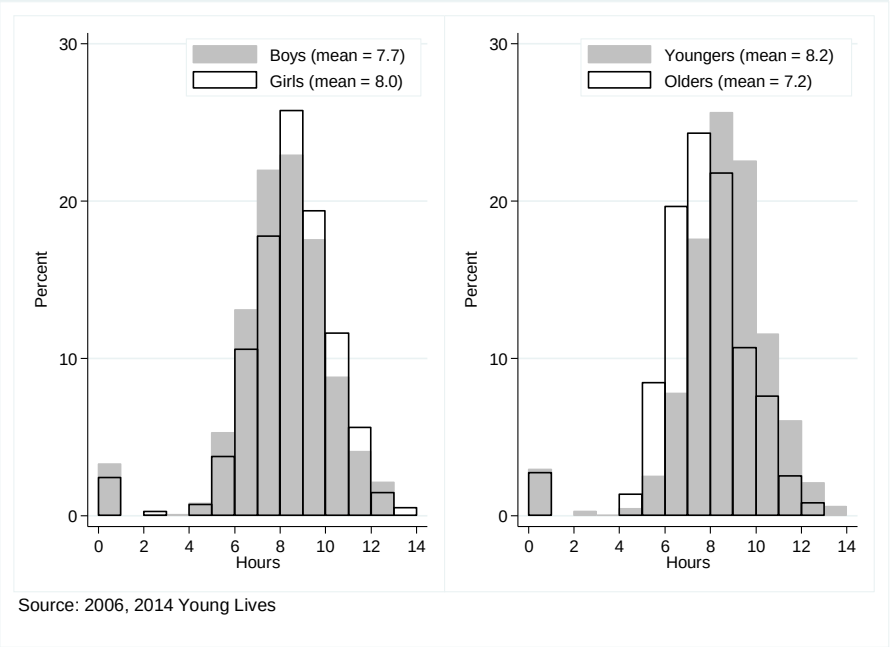

Figure 3.1. Histogram of study hours

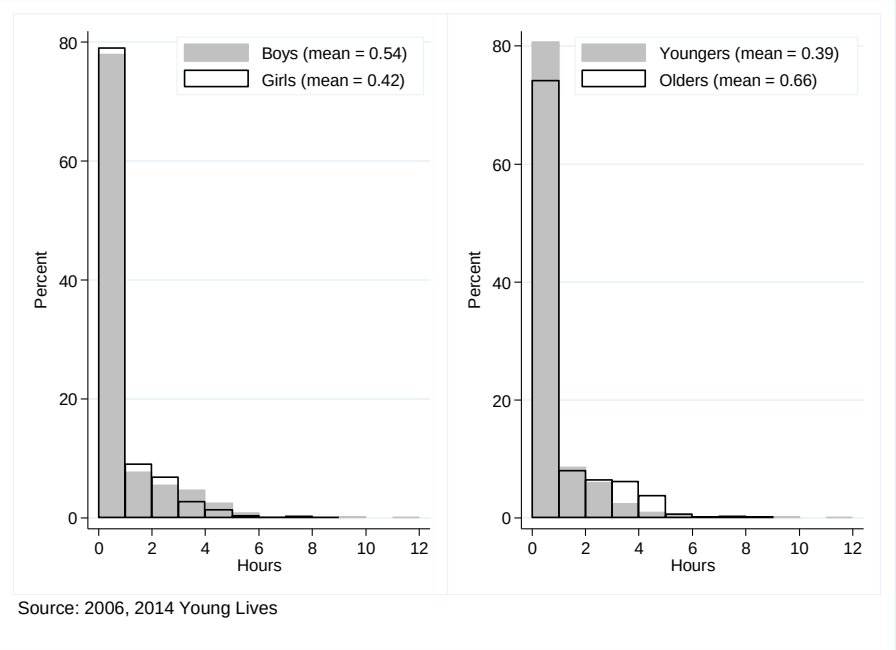

Figure 3.2. Histogram of work hours

The key explanatory variable is parental academic aspirations for a child. In the YLS, the question: 'Ideally what level of formal education would you like NAME to complete?' was answered by the main caregiver. Most of the main

caregivers are a biological parent of the child (95 per cent), so their answers are used to represent parental academic aspirations. The possible answers for this question include none, grades one to 12, post-secondary or vocational training, bachelor degree, and master degree. Figure 3.3 shows the distribution of parental aspirations by gender and cohort. In Figure 3.3, grades one to 12 are converted to the levels of education: primary education (grade 6), lower secondary education (grade 9), and upper secondary education (grade 12). The data indicate a high level of aspiration. Over 70 per cent of parents hope that their child will obtain a university degree and above when the child grows up; over 20 per cent of parents want their child to complete high school education or vocational training. On average, parental aspirations for girls are higher than for boys, and aspirations are quite similar for the two cohorts of children. In the econometric analysis, the level of education that parents want their child to achieve is converted into years of schooling.

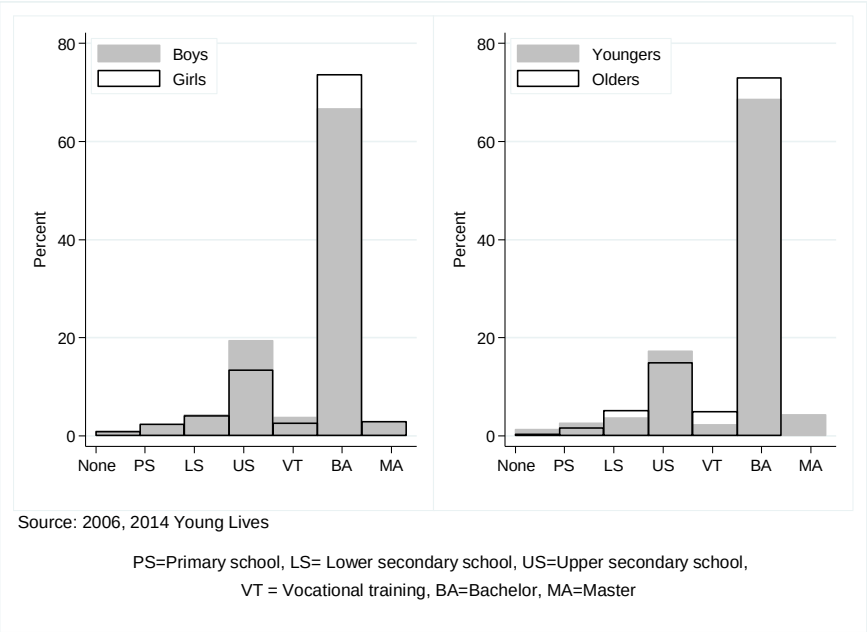

Figure 3.3. Histogram of parental academic aspirations for a child

Table 3.1. Descriptive statistics of variables

Variables	Mean	Standard deviation
Hours of study	7.84	2.12
Hours of economic work	0.49	1.17
Hours of housework	1.37	1.24
Parental aspirations	14.7	2.76
Younger Cohort	0.66	0.48
Male	0.51	0.5
Child's age in month	146.8	3.77
BMI-for-age z-score (-1)	-0.85	1.2
Birth order	1.85	1.06
Child's highest grade	5.62	0.92
Adjusted PPVT	0.12	22.72
Non-Kinh child	0.14	0.34
Number of children aged 0-5	0.26	0.51
Education of caregiver	7.15	3.94
Main caregiver is a parent	0.95	0.21
Wealth index	0.58	0.15
North East	0.2	0.4
Red River Delta	0.2	0.4
South Central Coast	0.2	0.4
Mekong River Delta	0.21	0.41
Ln(average local wage)	9.81	0.45
Number of children	2749	

Note: (-1) is the value in the previous survey; Data source: 2006 and 2014 YLS

Other explanatory variables include the child, household and commune characteristics. The descriptive statistics of these variables are summarized in Table 3.1. The Peabody Picture Vocabulary Test (PPVT) scores are controlled in the models because they can represent the Wechsler Intelligence Scale for Children (WISC) (Anderson and Flax 1968), and the Intelligence Quotient (IQ) (Altepeter 1989). It is noteworthy that in the analysis of work hours, both household income and expenditure are potentially endogenous in the models due to a mutual relationship between these economic variables and the child's work. Using the wealth index to represent the economic status of households helps to avoid the endogeneity problem to a certain extent. The wealth index is constructed from a housing index, an access to service index, and a consumer durable index, which economic benefit from the child's work is generally insufficient to contribute to.[4]

3.5.2. Endogeneity and instrumental variables

Parental academic aspirations for a child are likely to be endogenous in the models due to unobservable factors. For example, if a child is the favourite child among his or her siblings, parents are more likely to desire a higher level of education for that child and less likely to send that child to work. Additionally, parents who worked in their childhood tend to require their child to work as well as may have low aspirations for their child.

This study uses the average local wage relative to the education level that parents aspire their child to achieve as an instrumental variable for parental

[4] The definitions of variables are summarized in table A1 in Appendix.

aspirations[5]. When parents think about the level of education they desire for their child, one of the first things they consider is potential income that the child can obtain with that education level. For instance, parents might want their child to attain university degree because they observe the wealthy lives of university graduate people and hope that their child will be able to have that in the future. Alternatively, if parents find that income of high school graduates, on average, is good enough, parents might want their child to only finish high school. Specifically, the instrumental variable, the average local wage, is calculated for people who (1) are the same gender as the child, (2) are 15 to 30 years old, (3) have the highest education level that corresponds to the education level that parents aspire for their child and (4) live in the same economic region as the child. The age category of 15 to 30 is selected to construct the instrumental variable because these ages represent early career periods. The underlying argument is that parents mostly care about the beginning of their child's career. After the age of 30, their child is mature enough to be independent. Figure 3.4 shows a positive correlation between parental aspirations and the average local wage for both younger and older cohorts.

5. Data from the 2006 and 2014 Vietnam Household Living Standard Survey (2006 and 2014 VHLSS) are used to extract the instrumental variables. An individual's wage includes all receivables, bonus, and allowance in cash or in-kind converted to the monetary unit that he or she receives from all of the jobs. For more details of the 2006 and 2008 VHLSS, see Vietnam General Statistics Office (2008, 2016).

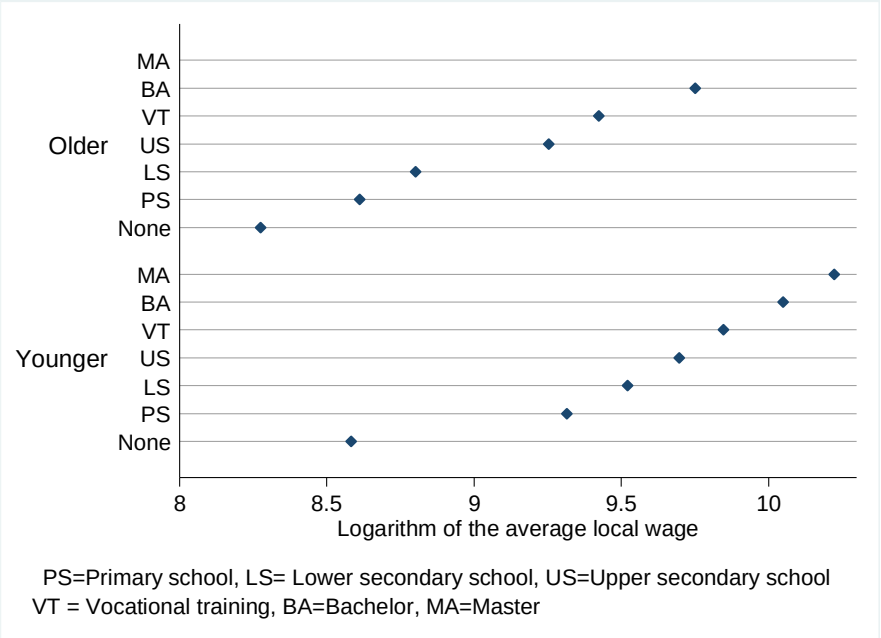

Figure 3.4. Parental aspirations by the average local wage and cohort

3.5.3. Model specification

Hours of an activity are positive if the child participates in the specified activity and 0 if the child is not involved in the activity. In this case, hours of the activity constitute a censored variable. However, the histograms show that only the child's economic work hours are censored and pile up at 0 (Figure 3.2). The hours of study appear to be normally distributed (Figure 3.1). For this reason, the Tobit model seems to be an appropriate specification for economic work hours and OLS still provides efficient estimators of study hours. Because a child's time is jointly allocated in a single decision-making process, hours of economic work and study should be simultaneously estimated in a single system of equations. Given the potential endogeneity of parental aspirations, an additional reduced form equation, apart from equations for study and work hours, is included in the system to estimate aspirations on instrumental and exogenous variables.

In summary, in the context of a child's time allocation and the potential endogeneity of parental aspirations, the empirical model simultaneously estimates a system of 2 linear equations for hours of study and parental aspirations, and 1 Tobit model for hours of work. This system is called a mixed process model (Roodman 2011). The specific model can be written as follows.

$$Y_1 = X\beta_1 + \phi_1 Y_3 + u_1 \tag{3.23}$$

$$Y_2 = \max(0, X\beta_2 + \phi_2 Y_3 + u_2) \tag{3.24}$$

$$Y_3 = X\alpha + Z\lambda + v \tag{3.25}$$

where Y_1 and Y_2 are observable hours of study and economic work, respectively; X are explanatory variables including the child, household and commune characteristics; Y_3 denotes parental academic aspirations for the child; Z represents the instrumental variables, i.e. the average local wage.

The error terms u_1, u_2 and v are assumed to follow a multivariate normal distribution. In particular,

$$\begin{pmatrix} u_1 \\ u_2 \\ v \end{pmatrix} \sim N\left(0, \begin{bmatrix} \delta_1^2 & \rho\delta_1\delta_2 & \gamma_1\delta_1\delta_v \\ \rho\delta_2\delta_1 & \delta_2^2 & \gamma_2\delta_2\delta_v \\ \gamma_1\delta_v\delta_1 & \gamma_2\delta_v\delta_2 & \delta_v^2 \end{bmatrix}\right)$$

where δ_1^2, δ_2^2 and δ_v^2 are variances of u_1, u_2 and v, respectively; ρ is the correlation coefficients between u_1 and u_2; γ_i is the correlation coefficient between u_i and v.

3.6. Estimation results

3.6.1. Impacts of parental academic aspirations on a child's hours of economic work and study

This section discusses the estimation results for the system of equations (3.23) – (3.25), as shown in Table 3.2. The discussion is first about the equations of the main interest, hours of work and study, followed by the equation of parental academic aspirations.

The estimation results are in line with those derived from the theoretical model: children of parents with higher academic aspirations spend more time studying and less time working than their counterparts. As expected, parental academic aspirations have a larger impact on study than on work due to the direct link between the aspirations and the child's study. The point instrumented estimates show that a 1 year increase in parental aspirations for the child's schooling, on average, leads to a rise of 0.2 hours in study time and a decrease of 0.12 hours in work time. The effects are statistically significant at 1 per cent level in both equations.

Impacts of other explanatory variables on work hours are quite different, even opposite, to those on study hours, due to the trade-off between the two activities. Some notable results are follows. Children in the younger cohort study more and work less than their counterparts in the older cohort. While there is no difference in work hours between boys and girls, girls significantly spend more time studying than boys do. Minority children on average have to work 1 hour more than their majority counterparts. This also reflects the fact that minority children suffer from

more disadvantages than the Kinh ethnic children. Number of children aged 0-5 living in the family significantly reduces total hours of economic work for a child. One possible reason is that the 11-13 year old children might spend time taking care of their younger siblings instead of doing economic work.

Other variables, such as BMI-for-age z-score, birth order, PPVT and wealth index do not significantly affect the child's time allocation. Related to geographic factor, children in the city region, which is the base region in the models, work least and study more than children in most of the other region, except for those in Red River Delta.

Table 3.2. Mixed process models: hours of study, economic work and parental aspirations

	Hours of study	Hours of economic work	Parental aspirations
Parental aspirations	0.199***	-0.119***	
	(0.021)	(0.046)	
Younger Cohort	0.831***	-0.511***	-2.548***
	(0.079)	(0.183)	(0.082)
Male	-0.232***	0.164	-1.212***
	(0.066)	(0.156)	(0.066)
Child's age in month	-0.054***	0.025	-0.032***
	(0.010)	(0.023)	(0.009)
BMI-for-age z-score (-1)	0.001	-0.040	0.094***
	(0.029)	(0.074)	(0.028)
Birth order	-0.032	0.161**	-0.019
	(0.033)	(0.072)	(0.031)
Child's highest grade	0.537***	-0.168*	0.417***

	Hours of study	Hours of economic work	Parental aspirations
	(0.047)	(0.098)	(0.042)
Adjusted PPVT	0.001	-0.005	0.001
	(0.002)	(0.004)	(0.002)
Non-Kinh child	-0.091	0.958***	0.166
	(0.139)	(0.289)	(0.132)
Number of children aged 0-5	-0.074	-0.443***	-0.091
	(0.066)	(0.157)	(0.063)
Education of caregiver	0.032***	-0.051*	0.061***
	(0.011)	(0.029)	(0.011)
Main caregiver is a parent	0.151	0.842**	0.155
	(0.156)	(0.426)	(0.148)
Wealth index	0.548*	-0.614	2.961***
	(0.317)	(0.751)	(0.292)
North East	-0.705***	2.963***	3.451***
	(0.136)	(0.349)	(0.141)
Red River Delta	0.361***	1.567***	3.624***
	(0.109)	(0.312)	(0.128)
South Central Coast	-0.597***	1.822***	3.040***
	(0.117)	(0.317)	(0.124)
Mekong River Delta	-0.583***	0.969***	4.471***
	(0.119)	(0.336)	(0.140)
Ln(average local wage)			5.792***
			(0.117)

Notes: Data source: 2006 and 2014 YLS; Significant level: *10%; **5%; ***1%;
Standard errors are in parentheses;

Estimation results for parental aspirations, equation (3.25), are presented in Table 3.2, Column 3. The coefficient of the instrumental variable, the average local wage, is significant and has an expected sign. In particular, a 1 per cent rise in the average local wage is associated with a 0.06 year increase in parental aspirations. In addition, parental aspirations are higher for girls and children in the older cohort than for boys and those in the younger cohort, respectively. Education of parents and economic condition are positively associated with parental aspirations for the child.

3.6.2. Impact of parental academic aspirations on a child's hours of housework

In this section, apart from the 2 main activities of interest in the literature: economic work and study, housework is considered as another activity of children, and thus examined under the impact of parental aspirations. Hours of housework consist of the time that a child does chores and takes care of younger children. Around 75 per cent of children are involved in housework; about 15 per cent of them devoted 3 hours or more per day to such activity (Figure 3.5). Hours of housework appear to be normally distributed, so they can be estimated as a linear function of explanatory variables. In terms of modelling, hours of housework can be represented by equation (3.26).

$$Y_4 = X\beta_4 + \phi_4 Y_3 + u_4 \tag{3.26}$$

where Y_4 is the time that a child does housework; the error terms u_4 are assumed to follow a multivariate normal distribution.

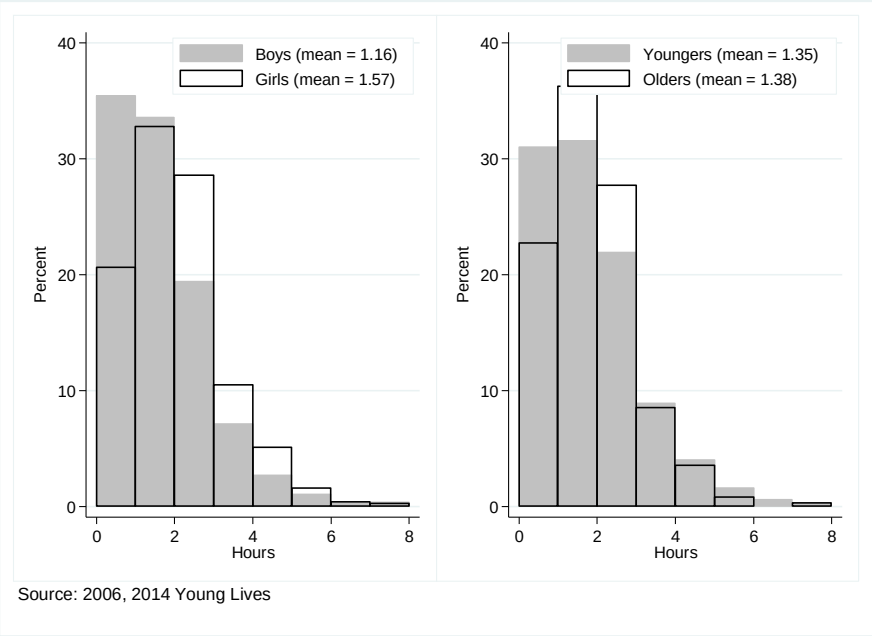

Figure 3.5. Histogram of housework hours

Table 3.3. Mixed process models: hours of study, economic work and housework, and parental aspirations

	Hours of study	Hours of economic work	Hours of housework	Parental aspirations
Parental aspirations	0.199***	-0.117**	-0.023	
	(0.021)	(0.046)	(0.018)	
Ln(average local wage)				5.792***
				(0.117)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses; Control variables: Child characteristics (gender, age, BMI score, birth order, highest grade, PPVT, ethnicity), household characteristics (number of 0-5 year old children, caregiver's education, caregiver is parent, wealth index), and regional dummy variables; The full estimation results are presented in Appendix, Table A2.

The empirical model is, now, extended to include equations (3.23) – (3.26). A summary of estimation results is provided in Table 3.3. The results suggest that

parental academic aspirations for a child insignificantly affect the time allocated to housework. Estimation results for study, work and parental aspirations are almost unchanged compared to those in Table 3.2.

3.6.3. Gender difference in the impact of parental academic aspirations

Table 3.4. Mixed process models: hours of study, economic work and housework, and parental aspirations, by gender

	Hours of study	Hours of economic work	Hours of housework	Parental aspirations
		Boys		
Parental aspirations	0.172***	-0.157**	-0.003	
	(0.032)	(0.070)	(0.029)	
Ln(average local wage)				5.731***
				(0.181)
		Girls		
Parental aspirations	0.241***	-0.142***	-0.043**	
	(0.027)	(0.054)	(0.021)	
Ln(average local wage)				7.031***
				(0.150)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses; Control variables: Child characteristics (age, BMI score, birth order, highest grade, PPVT, ethnicity), household characteristics (number of 0-5 year old children, caregiver's education, caregiver is parent, wealth index), and regional dummy variables; The full estimation results are presented in Appendix, Table A3 and Table A4.

This section estimates the models of a child's time allocation for boys and girls separately to explore the gender perspective in the impacts of parental aspirations.

The estimation results are presented in Table 3.4. As shown in the estimation for boys (Panel A), aspirations of parents still exert an increasing impact on the time that a child spends studying, and a reducing effect on hours of work, the effect is, however, more pronounced for studying hours spent by girls and working hours committed by boys. Interestingly, parental aspirations now impose a five per cent level significant impact on hours of housework for girls (even still insignificant for boys). Specifically, a one-year increase in parental academic aspirations leads to a decrease of 0.04 hours that girls do housework every day.

3.6.4. Robustness of the estimated impacts

Although an advantage of the mixed process model (3.23) – (3.26) is to estimate coefficients based on the accurate distribution the of the dependent variable in each equation, it might be criticized for providing inefficient and even inconsistent estimators in the case of heteroscedasticity (Roodman 2011). Therefore, in this section, the generalized method of moments (GMM) is applied to estimate the system of equations describing the child's time allocation. By calculating the optimal weighting matrix, GMM provides consistent and efficient estimators in the appearance of endogeneity and heteroscedasticity (Wooldridge 2010). It is noteworthy that GMM is, however, applicable only for general linear models. As a result, hours of activities carried out by the child and parental aspirations are assumed to be linear functions of control variables and instrumental variables in the GMM estimation.

The main results of the GMM estimation are presented in Table 3.5. It can be seen that the effects of parental aspirations derived from GMM estimation are

consistent with those from the mixed process model. Specifically, parental aspirations increase studying time and decrease work hours. Although parental aspirations have no effect on hours of housework for boys, they significantly reduce the time that girls do chores and take care of younger siblings. The magnitude of effects is almost unchanged for hours of study and housework, but under the assumption that hours of economic work are normally distributed, the impact of parental aspirations on this activity is underestimated.

Table 3.5. GMM estimation: Effects of parental aspirations on hours of study, economic work and housework

	Hours of study	Hours of economic work	Hours of housework
	All children		
Model for all children	0.199***	-0.080***	-0.030*
	(0.027)	(0.020)	(0.016)
	Boys		
Model for boys	0.172***	-0.105***	-0.014
	(0.037)	(0.031)	(0.022)
	Girls		
Model for girls	0.241***	-0.074***	-0.050**
	(0.038)	(0.022)	(0.021)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses; Control variables: Child characteristics (gender, age, BMI score, birth order, highest grade, PPVT, ethnicity), household characteristics (number of 0-5 year old children, caregiver's education, caregiver is parent, wealth index), and regional dummy variables. Instrumental variables: logarithm of the average local wage; The full estimation results are presented in Appendix, Table A5, Table A6 and Table A7.

3.6.5. Impacts of hours of study and work on education outcomes

Although the impacts of a child's time allocation on education outcomes are not the focus of this paper, as mentioned above, it is interesting to look at the completed picture of the relationship between parental academic aspirations, child's time allocation and academic achievement. Hence, this section establishes a completed model, in which parental academic aspirations influence the allocation of a child's time, which, in turn, affects the child's academic performance. Academic performance is represented by maths test scores, which is scaled from 0 to 10. The distribution of maths test scores is illustrated in Figure 3.6.

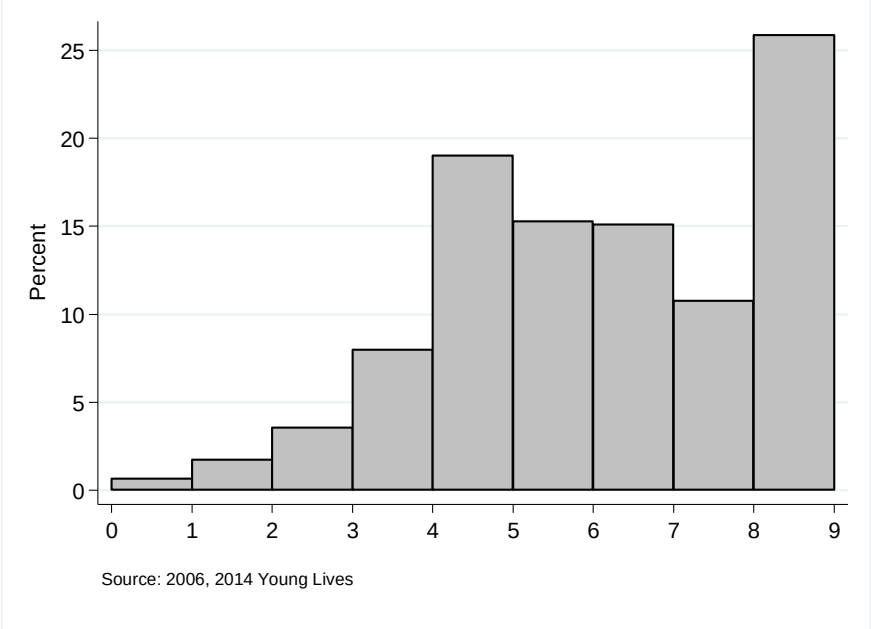

Figure 3.6. Histogram of Maths test scores

Since the test scores are relatively normally distributed, they can be estimated with a linear function, as shown equation (3.27).

$$Y_5 = X\beta_5 + \alpha_1 Y_1 + \alpha_2 Y_2 + u_5 \quad (3.27)$$

Where, Y_5 is maths test scores; the error terms u_5 are assumed to follow a multivariate normal distribution.

Table 3.6. Mixed process models: Maths scores, hours of study, economic work, and parental aspirations

	Maths score	Hours of study	Hours of economic work	Parental aspirations
Hours of study	0.842***			
	(0.116)			
Hours of economic work	-0.118**			
	(0.057)			
Parental aspirations		0.209***	-0.120***	
		(0.021)	(0.045)	
Ln(average local wage)				5.816***
				(0.114)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses; Control variables: Child characteristics (gender, age, BMI score, birth order, highest grade, ethnicity), household characteristics (number of 0-5 year old children, caregiver's education, caregiver is parent, wealth index), and regional dummy variables. Instrumental variables: logarithm of the average local wage; The full estimation results are presented in Appendix, Table A8.

The empirical model is, now, extended to include equations (3.23), (3.24), (3.25) and (3.27). As indicated in Table 3.6, column 1, both hours of study and hours of work significantly affect the maths test scores and study time creates a larger impact in terms of absolute value. On average, one additional hour of study leads to a 0.8 point increase in the test scores, while working one more hour reduces the test scores by 0.1 point. The impacts of parental academic aspirations on the

time that a child spends studying and working are still similar to those found in section 3.6.1.

3.7. Summary of main findings and direction for future studies

This paper has investigated the impact of parental academic aspirations for a child on the child's time allocation. First, a theoretical model was outlined by extending the household utility maximization problem. In the model, parental aspirations are an augmenting factor that helps to increase the productivity of the child's study time. A statistical analysis of optimal time allocation shows that parental aspirations are positively associated with studying time and negatively associated with the working time of the child.

Then, the impact of parental aspirations was tested by econometric methods. Taking into account potential endogeneity, parental aspirations were instrumented by the average local wage of people having the same education level to that parents want their child to achieve. The empirical findings support the theoretical analysis. In particular, an increase in parental academic aspirations of 1 year, on average, leads to a decrease of 0.2 hours in working time committed by the child and a rise in study time of over 0.1 hours. Parental aspirations, despite being insignificant in the equation of housework for boys, still help to reduce hours that girls do housework. A further analysis shows that academic performance of a child is negatively associated with number hours of work, and positively associated with the time he or she spends studying.

The consistent findings relating to the role of parental aspirations in determining a child's studying and working time from both theoretical and empirical analyses suggest that in a fight again child labour and an attempt to increase educational attainment of the young generations, interventions or policies that raise parents' academic expectations and aspirations for a child can positively affect the child outcomes. The interventions can be those concerning to economic benefits of education, social returns to education, and changes in the evaluation of education. These findings are particularly important for a developing country like Vietnam, where child labour and dropouts are still common.

In regard to research related to this topic, the findings also show that the allocation of the child's time is a missing link in the relationship between parental aspirations and the child's academic achievements. Thus, future studies might consider the allocation of a child's time as a channel through which parental aspirations affect a child's educational attainment.

4. Ethnic gaps in child education outcomes in Vietnam

4.1. Introduction

Like poverty, inequality is multi-dimensional. In countries with multiple ethnic groups, understanding inequality across groups becomes crucial from an analytical as well as a policy perspective. In Vietnam, there are 54 different ethnic groups, of which the Kinh majority accounts for 86 per cent of the population; the share of all other ethnic groups is 14 per cent. The largest minority ethnic groups, such as Tay, Thai, Muong and Khmer, account for less than 2 per cent. The living standards of minority groups are much lower than those of the Kinh. For example, the poverty rate in 2010 was 12.9 per cent for the Kinh while more than 66 per cent of the minority population was poor (Badiani et al. 2013). In 2012, the per capita income of Kinh households was VND 23 million, which was more than double that of non-Kinh households (McCaig, Benjamin, and Brandt 2015). The share of households among the Kinh having permanent houses, safe water or hygienic toilet facilities is double that of the non-Kinh (GSO 2010b).

Ethnic inequality in education is also considerable. The results of the 2009 Vietnam population and housing census (GSO 2010b) show that the literacy rate of the population aged 10 and over is 96 per cent for Kinh compared to 78 per cent for minority groups. Differences in enrolment rates at primary, lower secondary, higher secondary schools and university between Kinh and non-Kinh children were 8 per cent, 26 per cent, 35 per cent and 18 per cent, respectively. The ethnic gap in education attainment is also noteworthy. The proportion of non-Kinh population

aged 15 and over with no schooling is 23 per cent compared to 3 per cent for the Kinh. The dropout rate for the non-Kinh is double that for the Kinh and the late enrolment rate is 5-times greater (WB 2009). The ethnic gaps in reading and maths test scores were found for all students aged 9 to 20 (Dang 2012). Years of schooling for minority ethnic people aged 15-25 were persistently lower than those for the majority counterparts during the period 1992-2014 (Dang and Glewwe 2017). In addition, educational inequality within a non-Kinh ethnic group is also high. The education Gini index for the Kinh population is 0.25 while for most other ethnicities the index ranges from 0.28 to 0.7 (Rew 2008).

Glewwe, Chen, and Katare (2015) and Arouri, Ben, and Nguyen (2016) are among the few to investigate the ethnic gap in education in Vietnam. Using the 2006 Young Lives survey (YLS) data, Glewwe, Chen, and Katare (2015) concluded that Kinh children had better reading and maths skills than non-Kinh children. Language barriers were an obstacle to minority children catching up with their majority peers. Blinder-Oaxaca decomposition results revealed that household expenditure and parents' education were the main contributors to the ethnic gaps in test scores. Applying the same methodology on pooled data from the 2006 YLS and 2009 YLS, Arouri, Ben, and Nguyen (2016) pointed out that child health, mother's education and household demographic factors were mainly responsible for the difference in education outcomes between Kinh and non-Kinh children. In research on ethnic earnings inequality in Vietnam (Baulch et al. 2010; Doan 2011; Imai, Gaiha, and Kang 2011; Pham and Reilly 2009; Van de Walle and Gunewardena 2001), low returns to education in non-Kinh groups compared to the Kinh group were also documented. Low returns to education might perpetuate the ethnic gap in

education because such returns discourage investment in education among minority populations. Other possible causes for the non-Kinh's lagging performance in education are difficulties in physical access to school and the quality of schools in minority areas (WB 2009). Most minority ethnic groups are located in remote and mountainous areas, where infrastructure is still limited.

Ethnic and racial disparities in education have been found in both developed and developing countries. Cook and Evans (2000) focused attention on the convergence of reading and maths test scores between 13-year-old black and white students in the United States during the period 1970-1988. They decomposed the differences in test scores into variation in school and family characteristics and within-school changes. Their findings indicated that three quarters of the convergence was attributable to within-school changes while the variation in family backgrounds and school characteristics accounted for the rest. The changes in the quality of schools negligibly influenced differences in maths test scores while they considerably reduced the divergence in reading test scores during the period. In spite of this convergence, persistent gaps in education attainment and the dropout rates between young black and white Americans still exit and the gaps are even wider if the prison population is counted (Ewert, Sykes, and Pettit 2014). In Australia, indigenous/non-indigenous gaps in reading, writing and numeracy test scores were found in all states for all grade 3, grade 5 and grade 9 students (Ford 2013). The gaps even occurred before children went to school, as found by Leigh and Gong (2009) who examined the cognitive test scores for 4 and 5-year-old children. Socioeconomic differences between indigenous and non-indigenous populations mostly explained the gaps. However, Baert and Cockx (2013)

discovered that the unexplained part became considerable if schooling delays were taken into account. Moreover, the authors pointed out that the education attainment gap between the third generation non-Western populations in Belgium and native Belgians started to rise in year 4 of secondary school. Sakellariou (2008) sought an explanation for the test score gap between indigenous and non-indigenous students in Peru in 1997. The results implied that the peer effect, measured by the share of non-indigenous students and the average parents' education of students in the class, explained between one half to two thirds of the gaps. School quality was found to be unimportant for the test score gaps.

This paper aims to discover the key contributors to the education gap between Kinh and non-Kinh children in Vietnam. The paper contributes to the literature by investigating the ethnic gap across different education outcomes. It is hypothesised that determinants vary in enrolment, schooling progress and performance, and in different ethnic groups, thus leading to variation in their contribution to the ethnic gaps in the three education outcomes, which has not been explicitly specified in the aforementioned studies in education in Vietnam. In addition, the current study, unlike previous empirical studies which used household data to estimate test scores, employs school data to capture peer, class and school characteristics in models of test scores. Finally, various econometric techniques including decomposition for a multilevel model are used to deal with different measurements of education outcomes. Although multilevel models have been widely applied to analyse school survey data, to the best of my knowledge multilevel model decomposition techniques have not yet been used in the literature to examine education outcomes.

4.2. Data

This paper employs data from two different surveys, the 2009 Young Lives Survey (YLS) and the 2011-2012 Young Lives School Survey (YLSS). The former provides information related to enrolment and schooling progress and the latter supplies data on performance at school in Vietnam.

The YLS tracks data on two groups of children who were born in 2001-2002 (the younger cohort) and in 1994-1995 (the older cohort) over a 15-year period. In Vietnam 2000 children in the younger cohort and 1000 children in the older cohort were selected in the sample. The sampling procedures were designed to ensure that the sample proportionally covered urban, rural and mountainous areas, in the northern, central and southern regions in Vietnam.

According to Nguyen (2008), the YLS, however, lacks representativeness because the survey was designed to focus on poor children and was based on non-random sampling. The poverty indexes and access to basic services in the YLS are lower than those in national representative surveys (Nguyen 2008). Table 4.1 displays summary statistics concerning education and ethnicity calculated from YLS and two other nationally representative samples, the 15 per cent sample of 2009 Vietnam Population and Housing Census (MPC 2017), and 2010 Vietnam Household Living Standard Survey. It can be seen that the YLS produces relatively similar enrolment rates to the survey conducted in the same year, the 2009 Census, except for the enrolment rate of non-Kinh children in the older cohort, which is 12 per cent lower.

Table 4.1. The share of ethnic minorities in population and the gross enrolment rates

	2009 Young Lives	2009 Census	2010 VHLSS
Non-Kinh population (%)	13.8	14.3	15.7
Enrolment rate (%)			
Younger cohort (7-8)	98.5	96.4	99.5
Kinh	99.7	98.4	99.7
Non-Kinh	91.4	91.1	98.5
Older cohort (14-15)	76.2	76.7	80.1
Kinh	80.0	81.7	83.6
Non-Kinh	50.6	62.9	69.2

Data source: 2009 YLS, 15 per cent sample of Vietnam Population and Housing Census, 2010 VHLSS

The difference in the enrolment rates of the older cohort between the surveys can be partly explained by the difference in the time that the surveys were conducted. In particular, the census date was 1st April 2009, which fell in the second semester of the 2008-09 school year, when most 15-year-old children in Vietnam were in the last year of lower secondary school. The 2009 YLS data were collected from September to December, 2009, the first semester of the 2009-10 school year, when most 15-year-old children were in the first year of upper secondary school if they were enrolled in school. Hence, in between the two surveys there was a transition from lower secondary school to upper secondary school, in which a number of students, especially disadvantaged students, might drop out of school. To take this transition into account in the sample of the 2009 census, it would be more precise to look at children aged 15-16, rather than those aged 14-15. The enrolment rate of non-Kinh children aged 15-16 in the sample of 2009 census is 51

per cent, almost the same as that of the older cohort in the 2009 YLS (see Table 4.1).

Therefore, data from the YLS still seem to be valid for examining the ethnic gap in education. Moreover, according to Nguyen (2008), YLS data, despite not being suitable for constructing indicators related to children's welfare, are useful for modelling and analysing causal relations. After dropping observations with missing information, the final sample includes 2912 children. Of the total sample, 971 children belong to the older cohort, who were 14 or 15 years old at the time of the survey, and 1950 children are from the younger cohort, 7 or 8-year-old children (see Table 4.2). The share of minority ethnic children is 13.8 per cent in the YLS.

Table 4.2. Sampling of 2009YLS and 2011-2012YLSS

	Ethnicity		Gender		Total
	Kinh	Non-Kinh	Boys	Girls	
2009 YLS	2516	405	1483	1438	2921
Younger cohort (7-8)	1,671	279	1,004	946	1950
Older cohort (14-15)	845	126	481	490	971
2011-2012 YLSS	2,879	399	1,734	1,550	3,284

Data source: 2009 YLS and 2011-2012 YLSS

The YLSS was conducted by the Young Lives project in Vietnam in 2011-2012. The aim of the survey is to collect information about students' backgrounds and their learning outcomes. The sample contains the Young Lives children, who were in the younger cohort and enrolled in grade 5 in the school year 2011-2012, and their peers. In the class in which the Young Lives children were enrolled, their

peers were randomly selected so that the maximum number of students selected in each class was 20. The sample contains 3284 grade 5 students, of which 1138 are Young Lives children from 176 classes and 92 school sites. Data collection was implemented at child, teacher and principal levels. The share of non-Kinh children in the YLSS is 12.2 per cent, lower than that in the YLS, which speaks to the lower enrolment rates of ethnic minority groups. After dropping observations with missing information, the sample of the YLSS used in analysis includes 3218 children (see Table 4.2).

4.3. Methodology

4.3.1. Multi-level model: estimation and decomposition

School survey data on students' performance can generate three levels of data: individual, class and school. The multilevel structure of the data creates dependencies among the levels. For example, students studying in one class share the same class, teacher and peer characteristics. Therefore, the performance of one student in the class is not independent from that of other students. The consequence of ignoring the dependencies is to underestimate standard errors, which leads to finding significant impacts when they do not exist (Rasbash 2008).

A multilevel model of students' performance follows.

$$S_{ijk} = \alpha + \gamma E_{ijk} + \beta_1 X_{1ijk} + \beta_2 X_{2jk} + \beta_3 X_{3k} + v_k + u_{jk} + \varepsilon_{ijk} \quad (4.1)$$

where S is scores and X_1, X_2, X_3 comprise explanatory variables. For the sake of decomposing the ethnic differences in education performance, a dummy variable,

E, defining minor ethnicity is included in the model. Subscripts *i, j, k* indicate data at individual, class and school levels, so X_1, X_2, X_3 are vectors of explanatory variables at individual, class and school levels, respectively. Similarly, the error terms in the multi-level model are split into three components: v_k are school level error terms, representing school random effects; ν_{jk} are class level error terms, representing class random effects; and ε_{ijk} are individual level error terms.

The estimated equation (1) is

$$S_{ijk} = \hat{\alpha} + \hat{\gamma}E_{ijk} + \hat{\beta}_1 X_{1ijk} + \hat{\beta}_2 X_{2jk} + \hat{\beta}_3 X_{3k} + \hat{v}_k + \hat{u}_{jk} + \hat{\varepsilon}_{ijk} \tag{4.2}$$

The difference in the means of the test scores between the minority ethnic group, *m*, and the majority ethnic group, *M*, can be decomposed by using the method proposed by Jacobson, Robinson, and Bluthenthal (2007), as follows.

$$\Delta^{multilevel} = \overline{S}^M - \overline{S}^m \tag{4.3}$$

$$\Delta^{multilevel} = \sum_{i=1}^{3} \hat{\beta}_i \left(\overline{X}_i^M - \overline{X}_i^m\right) + \left(\overline{\hat{v}}^M - \overline{\hat{v}}^m\right) + \left(\overline{\hat{u}}^M - \overline{\hat{u}}^m\right) - \hat{\gamma} \tag{4.4}$$

In equation (4.4), $\sum_{i=1}^{3} \hat{\beta}_i \left(\overline{X}_i^M - \overline{X}_i^m\right)$ is the explained part, providing the contribution of explanatory variables to the ethnic gap in the test scores and $\left(\overline{\hat{v}}^M - \overline{\hat{v}}^m\right) + \left(\overline{\hat{u}}^M - \overline{\hat{u}}^m\right)$ is the random part generated by school and class level errors; $\hat{\gamma}$ is the unexplained component created by unobservable factors.

4.3.2. Probit model: estimation and decomposition

Consider a Probit model of an education outcome measured by a binomial variable, *Y*, e.g. school enrolment:

$$P(Y=1|X)=F(X\beta) \tag{4.5}$$

where Y = 1 if the child enrols in school, and Y = 0 otherwise. The right-hand side of equation (4.5) is the conditional probability that Y = 1. F is a cumulative normal distribution function. Equation (4.5) can be estimated by maximum likelihood methods.

The decomposition of the Probit model was developed by Powers, Yoshioka, and Yun (2011) as follows.

$$\Delta^{Probit} = \overline{P}^{M} - \overline{P}^{m} = \overline{F(X^{M}\beta^{M})} - \overline{F(X^{m}\beta^{m})} \tag{4.6}$$

$$\Delta^{Probit} = \left\{\overline{F(X^{M}\beta^{M})} - \overline{F(X^{m}\beta^{M})}\right\} + \left\{\overline{F(X^{m}\beta^{M})} - \overline{F(X^{m}\beta^{m})}\right\} \tag{4.7}$$

The first component in the right-hand side of equation (4.7) is the explained part representing the difference in education outcomes attributable to differences in covariates, and the second component is the unexplained part caused by differences in estimated coefficients.

4.3.3. Variable selection

The descriptive statistics of variables used in the empirical analysis are shown in Table 4.3.

Education outcomes are measured by enrolment, schooling for age and test scores. The enrolment rate and the schooling for age index are calculated from the YLS while test scores are derived from the YLSS.

Table 4.3. Mean of variables in the models by ethnic groups and T_test of equal means

Variables	Kinh	Non-Kinh	Difference
From the YLS			
School enrolment	0.932	0.788	0.144***
Schooling for age index	0.959	0.766	0.193***
Schooling for age index = 1	0.881	0.59	0.291***
Age in month	125.449	122.252	3.197
Younger cohort	0.664	0.689	-0.025
Boy	0.505	0.523	-0.018
Height-for-age z score	-1.072	-2.079	1.007***
Health problems	0.072	0.104	-0.032**
Mother's year of schooling	6.528	1.272	5.257***
Father's years of schooling	6.845	2.101	4.744***
Number of siblings	0.829	1.432	-0.603***
Male head	0.866	0.936	-0.07***
Asset index	0.571	0.329	0.242***
Health shock	0.242	0.237	0.005
Newborn baby shock	0.054	0.109	-0.055***
School travel time	16.026	24.048	-8.022***
Paved road in commune	0.935	0.696	0.239***
Factory in commune	0.61	0.328	0.281***
Ln(population in commune)	9.175	8.412	0.763***
From the YLSS			
IRT-adjusted maths scores	18.861	14.398	4.463***

Variables	Kinh	Non-Kinh	Difference
IRT-adjusted Vietnamese scores	20.821	16.15	4.671***
Age in month	124.933	126.659	-1.725***
Boy	0.523	0.537	-0.015
Health problem	0.224	0.212	0.012
Speaking Vietnamese at home	0.985	0.38	0.605***
Mother's years of schooling	7.494	2.886	4.608***
Unknown mother's education	0.196	0.124	0.072***
Father's years of schooling	7.49	4.439	3.051***
Unknown father's education	0.239	0.15	0.089***
Older siblings	0.897	1.367	-0.47***
School travel time	11.496	16.525	-5.029***
Asset index	0.722	0.661	0.061***
Days absent from school of classmates	0.253	0.432	-0.18***
Grade repetition of classmates	0.039	0.07	-0.031***
Class size	30.566	18.698	11.868***
Television in classroom	0.098	0.003	0.096***
Dropout at school	0.043	0.204	-0.161***
School accesses to internet	0.428	0.148	0.280***
Newly established school	0.021	0.083	-0.062***

Notes: Data source: 2009 YLS and 2011-2012 YLSS; Significance level: *10%; **5%; ***1%

The enrolment rate of the younger cohort in the YLS is 98.5 per cent and that of the older cohort is 76.2 per cent. The ethnic gap in the enrolment rate for the former group of children is more than 8 per cent and that for the latter is almost 30 per cent. For all Young Lives children, the ethnic gap in the enrolment rate is 14.4 per cent (see Table 4.3).

The schooling for age index, denoted as SAGE (Ray and Lancaster 2005), is calculated by the formula

$$\text{SAGE} = \frac{\text{the highest grade attainted by the child}}{\text{the child's age - 6}} \qquad (4.8)$$

where 6 is the age that children in Vietnam start school. For 6-year-old children, who are currently studying grade 1, the index is replaced by 1. Hence, the values of SAGE range from 0 to 1 (see Figure 4.1). If a child starts school at age 6, does not repeat any grade and continues to enrol in school, her or his SAGE index is 1. Unlike enrolment, which describes the present status of schooling, SAGE takes into account late starts and grade repetition. Therefore, the SAGE index reflects any distortions in schooling progress (Dorman 2008). Although the age that children start primary school, as regulated by law, is 6, in some special cases children are allowed to start school later than the regulated age. This applies to children in remote areas, minority ethnic children, children migrating from abroad to Vietnam and children with disabilities. Among 2921 Young Lives children, 95 per cent of Kinh children started school at 6 or earlier compared to 75 per cent of non-Kinh children. The schooling for age index is also quite different for Kinh and non-Kinh children. The proportion of Kinh children with a SAGE index equal to 1 is 88 per cent and that of non-Kinh children is 59 per cent (see Table 4.4).

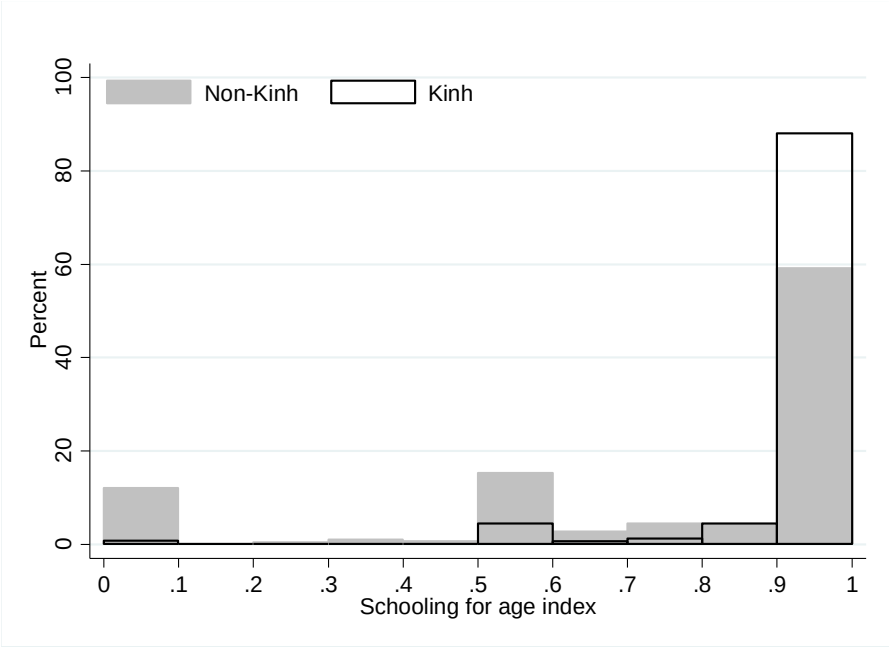

Figure 4.1. Histogram of Schooling for age index

In the YLSS, grade 5 pupils were requested to do tests in maths and Vietnamese. Each test contains 30 multiple choice questions with four options. A correct answer adds one mark to the raw test score. Because the tests were designed to assess students' understanding of the grade 5 curriculum, they did not include advanced questions, at which good students might be skilled (Rolleston et al. 2013). Moreover, the multiple choice format of the tests might lead to inflated test scores as a result of lucky guesses. Thus, Item Response Theory (IRT) (see (Baker 2001; Van Der Linden and Hambleton 1997) is used to adjust the test scores. For the sake of comparison, the IRT-adjusted test scores are normalised with the same means and variances to the raw scores. The IRT-adjusted test scores are highly correlated with the raw test scores, i.e. 0.989 for maths and 0.971 for Vietnamese. The distributions of raw and IRT-adjusted test scores for Kinh and non-Kinh pupils are shown in Figure 4.2 and Figure 4.3. The gaps exist in both maths and Vietnamese test scores and range from 4.4 points to 4.6 points (see Table 4.4).

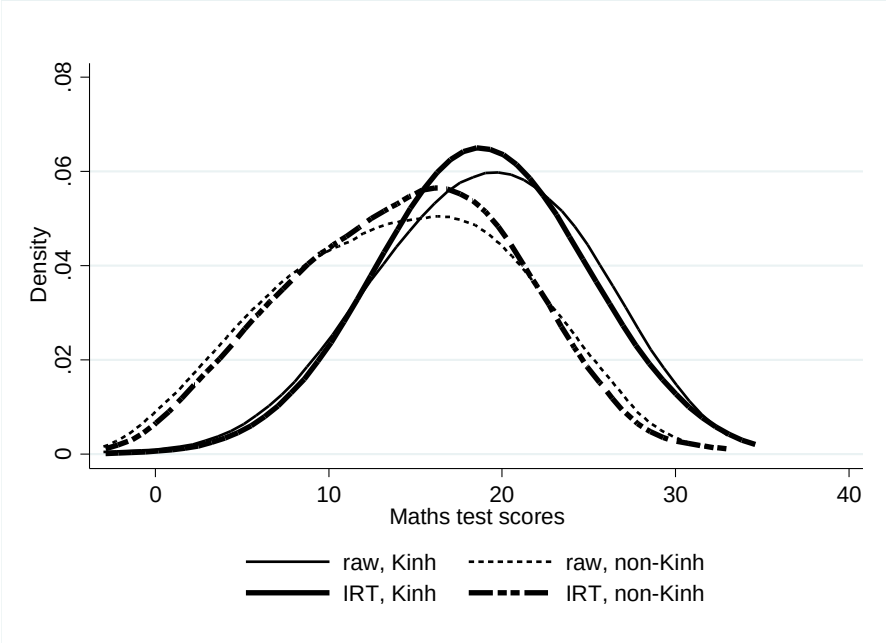

Figure 4.2. Distributions of raw and IRT-adjusted Maths test scores by ethnicity

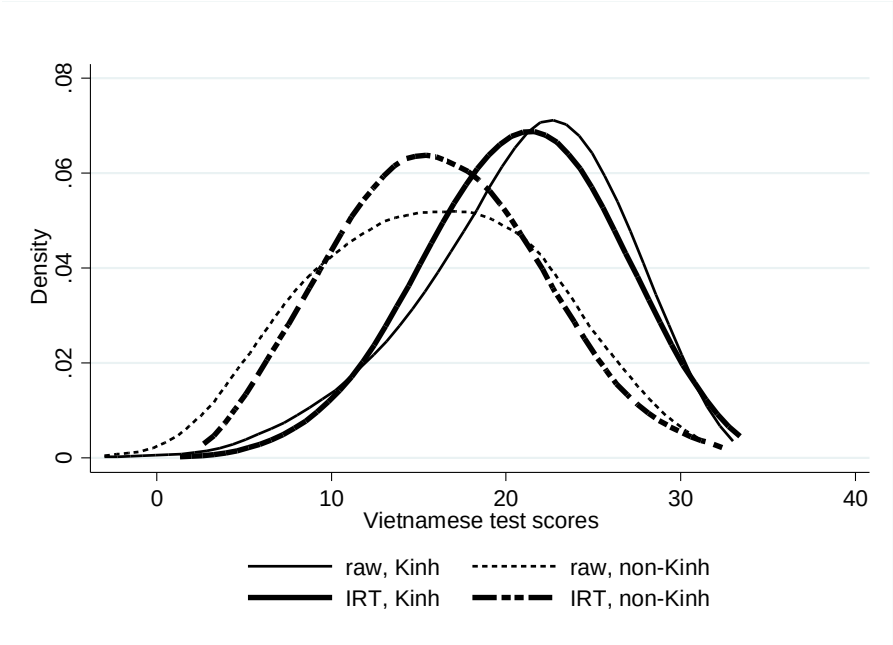

Figure 4.3. Distributions of raw and IRT-adjusted Vietnamese test scores by ethnicity

To make the estimation results comparable between the models of different education outcomes, explanatory variables in each model are constructed using the same method, where possible. In the estimations of enrolment and SAGE, explanatory variables consist of child, household and commune characteristics from YLS data. In the equation of test scores, the explanatory variables used are child, household, class and school characteristics from YLSS data. The common set of regressors among all models includes the child's age, a dummy for long-term health problems (except vision related problems due to the potential reverse impact of the child's studying on it), parents' years of schooling, the number of older siblings, school travel time and the asset index. The asset index, conducted by adopting the method used by YL (2002), represents the economic condition of a household, which is the simple average of 9 dummy variables for assets owned by the household: a television, radio, car, motorbike, bicycle, land phone, mobile phone, fan and computer[6].

In the school survey, most of the information about household characteristics was collected by interviewing students. One problem is that 20 per cent of grade 5 children did not know their parents' education level. In the models, parents' education is represented by years of schooling, which can be regarded as an interaction between years of schooling and a dummy representing the fact that

6. Using the asset index derived from the simple count method to represent household economic condition was found to yield consistent results as controlling for household expenditure (Montgomery et al. 2000), or using the index derived from the principal components analysis (Bollen, Glanville, and Stecklov 2002; Paxson and Schady 2007) as well as from various other methods (Filmer and Scott 2012).

parents' education is known by the child, and a dummy for 'Don't know' answers. Peer effects are controlled for by using information about classmates, i.e. the number of days absent from school and grade repetition. These variables are calculated by taking the mean values of interviewed children in the same class as the child in question.

4.4. Empirical results[7]

4.4.1. Enrolment

The Probit estimation for school enrolment[8] and the corresponding decomposition results are presented in Table 4.4.

4.4.1.1. Estimation results

In the Probit estimation for all children, the ethnicity of the child is statistically insignificant. The insignificance of ethnicity in the model does not imply the

7. The estimation results of the models with using the common set of regressors are provided in Table A9 and Table A10. Most results are consistent with those from the models with the full set of regressors.

8. An alternative method to estimate models of enrolment and SAGE by using household data is multi-level models, with child and commune level data. However, the nonlinear relationship between covariates and the depended variable leads to a difficulty in decomposition technique. Table A11 provides the estimation results of enrolment and SAGE from multilevel mixed effect Probit model. Because the estimated values of coefficients in the multilevel mixed effect Probit model are quite similar to those in Probit model (Table 4.4), I expect that the decomposition results derived from the two methods, if available, should be similar, too.

absence of the ethnic gap. This means that the ethnicity of a child does not significantly influent her (his) enrolment status when other variables are controlled for. The results show that girls and younger children are more likely to go to school than their counterparts. Most of the variables related to household characteristics are found to have a significant effect on enrolment. Parents' education and the asset index are positively associated with the probability of enrolment. In contrast, both the health shock and the newborn shock have negative effects on the enrolment rate. Additionally, children with more older siblings have less chance of participating in school. All variables representing commune characteristics have statistically insignificant impacts in the model.

When enrolment is estimated separately for Kinh and non-Kinh children, the results for Kinh children are almost the same as those interpreted above. For non-Kinh children, there are only three significant determinants: father's education, household economic condition and the newborn baby shock.

Table 4.4. Probit Models (marginal effects) and Decomposition Results for Enrolment

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Minority ethnic child	0.006				
	(0.014)				
Age in months	-0.002*	-0.002**	-0.003	0.006	-4.4
	(0.001)	(0.001)	(0.003)	(0.009)	
Younger cohort	0.025	-0.011	0.097	0.002	-1.2
	(0.090)	(0.099)	(0.290)	(0.006)	
Boy	-0.019**	-0.021***	0.001	0.000	-0.0
	(0.008)	(0.008)	(0.034)	(0.000)	
Height-for-age z score	0.003	0.005	0.005	-0.004	2.5
	(0.005)	(0.005)	(0.020)	(0.014)	
Health problem	-0.002	0.029	-0.079	-0.002	1.3
	(0.018)	(0.023)	(0.055)	(0.001)	

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Mother's year of schooling	0.005***	0.004***	0.005	-0.018	12.5
	(0.001)	(0.001)	(0.008)	(0.035)	
Father's years of schooling	0.005***	0.003**	0.021***	-0.072***	50.1
	(0.001)	(0.001)	(0.007)	(0.024)	
Older siblings	-0.012***	-0.011***	-0.012	-0.005	3.6
	(0.003)	(0.003)	(0.011)	(0.005)	
Male head	-0.008	-0.004	-0.021	-0.001	0.7
	(0.013)	(0.013)	(0.055)	(0.004)	
Asset index	0.224***	0.226***	0.289**	-0.050**	34.7
	(0.028)	(0.028)	(0.118)	(0.021)	
Health shock	-0.028***	-0.025***	-0.039	0.000	-0.1
	(0.009)	(0.009)	(0.037)	(0.000)	
Newborn baby shock	-0.042**	-0.025	-0.097**	-0.004*	2.6
	(0.017)	(0.021)	(0.047)	(0.002)	

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
School travel time	-0.000	-0.000	-0.000	-0.002	1.2
	(0.000)	(0.000)	(0.001)	(0.004)	
Paved road in commune	0.009	0.020	0.036	-0.006	4.3
	(0.014)	(0.016)	(0.049)	(0.009)	
Factory in commune	0.017*	0.007	0.056	-0.011	7.8
	(0.010)	(0.009)	(0.049)	(0.011)	
Ln(population in commune)	-0.001	0.013	-0.045	0.024	-16.9
	(0.009)	(0.009)	(0.043)	(0.024)	
Total				-0.143***	98.7
				(0.025)	
Mean predicted enrolment probability	0.912	0.932	0.788		
Pseudo R2	0.425	0.450	0.327		
Observations	2,921	2,516	405		

Notes: Data source: 2009 YLS; Significance level: *10%; **5%; ***1%; Standard errors in parentheses

4.4.1.2. Decomposition results

The decomposition result shows that 99 per cent of the ethnic gap is attributable to explanatory variables. Father's education accounts for half of the difference in enrolment rates between Kinh and non-Kinh children. The dominant role of father's education, which will be further explained in section 4.4.2.1, is partly due to its significant effects in the estimations for both Kinh and non-Kinh children as well as the large ethnic disparity in father's education (4.7 years of schooling). Likewise, the fact that the asset index in majority ethnic households is almost double that in minority households partly explains its contribution of 35 per cent to the ethnic enrolment difference. A further explanation for the role of household economic condition in determining children's education outcomes and the ethnic gaps will be provided in section 4.4.3.2. When factors are divided into three groups: the child, household and local characteristics, the child and local characteristics help to reduce the gap by 6 per cent and household characteristics contribute to the total ethnic enrolment gap of 105 per cent. Therefore, ethnic inequality in enrolment is mostly determined by ethnic differences in household characteristics, of which father's education and economic condition are two dominant contributors.

4.4.2. Schooling for age (SAGE)

Recall that the distribution of SAGE piles up at 1 (see Figure 4.1), so SAGE is a censored variable. Although the Tobit model is appropriate for analysing SAGE, there is a limitation in the relevant decomposition technique. The method of decomposition for Tobit models, developed by Bauer and Sinning (2010), only allows the decomposition of the total gap into an explained part and an unexplained

part in total. Thus, the contribution of individual variables, which is the main interest of this research, cannot be examined. For the sake of investigating the key determinants of the ethnic gap in education, estimation and decomposition for Probit models of the modified SAGE are used. In particular, the Probit models are employed to estimate the probability that a child has a SAGE index of 1. Estimation results are presented in Table 4.5.

4.4.2.1. Estimation results

Most of the estimation results for SAGE are consistent with what was found in the estimation for enrolment rates. The main differences follow. Ethnicity of the child is a significant determinant, implying that apart from explanatory variables in the model, unobservable factors also contribute to the ethnic differences in schooling for age. The height-for-age z score, despite having no impact on the child's present schooling status as shown in the estimation of enrolment, positively affects schooling progress. The height-for-age z score represents the child's health as a result of nutrition accumulation (WHO 1997). Thus, this might link to physical as well as mental health, and so influence a late start or grade repetition. For example, Haile et al. (2016) show that there is a positive association between the height-for-age z score and children's test scores.

Table 4.5. Probit models (marginal effects) and decomposition results for schooling for age

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Minority ethnic child	-0.048**				
	(0.020)				
Age in months	0.003	0.004**	-0.007	0.017	-5.7
	(0.002)	(0.002)	(0.005)	(0.012)	
Younger cohort	0.396***	0.494***	-0.283	-0.005	1.8
	(0.141)	(0.149)	(0.404)	(0.008)	
Boy	-0.031***	-0.028**	-0.026	-0.000	0.1
	(0.012)	(0.012)	(0.042)	(0.001)	
Height-for-age z score	0.024***	0.014**	0.116***	-0.090***	30.8
	(0.007)	(0.007)	(0.025)	(0.019)	
Health problem	-0.025	-0.006	-0.120*	-0.003*	1.0
	(0.022)	(0.024)	(0.066)	(0.002)	
Mother's year of schooling	0.005***	0.003**	0.021**	-0.084**	28.7

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.002)	(0.002)	(0.010)	(0.040)	
Father's years of schooling	0.006***	0.006***	0.005	-0.017	5.7
	(0.002)	(0.002)	(0.009)	(0.027)	
Older siblings	-0.010*	-0.011*	-0.018	-0.008	2.9
	(0.005)	(0.006)	(0.015)	(0.007)	
Male head	-0.021	-0.018	-0.031	-0.002	0.6
	(0.020)	(0.019)	(0.097)	(0.005)	
Asset index	0.287***	0.277***	0.330**	-0.061**	21.1
	(0.041)	(0.043)	(0.137)	(0.025)	
Health shock	-0.031**	-0.034***	0.006	-0.000	0.0
	(0.013)	(0.013)	(0.052)	(0.000)	
Newborn baby shock	-0.044*	-0.058**	-0.010	-0.000	0.1
	(0.024)	(0.023)	(0.070)	(0.003)	
School travel time	0.003***	0.005***	0.003***	0.019***	-6.5

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.001)	(0.001)	(0.001)	(0.006)	
Paved road in commune	0.057***	0.059**	0.023	-0.004	1.5
	(0.020)	(0.024)	(0.064)	(0.012)	
Factory in commune	-0.010	-0.027*	0.115*	-0.025*	8.5
	(0.014)	(0.014)	(0.064)	(0.014)	
Ln(population in commune)	0.025*	0.032**	0.079	-0.046	15.9
	(0.013)	(0.014)	(0.050)	(0.029)	
Total				-0.310***	106.5
				(0.027)	
Mean predicted SAGE	0.841	0.881	0.589		
Pseudo R2	0.223	0.171	0.238		
Observations	2,921	2,516	405		

Notes: Data source: 2009 YLS; Significance level: *10%; **5%; ***1%; Standard errors in parentheses

Another difference between the estimation results for enrolment and SAGE is school travel time. While time to travel from home to school has no influence on enrolment probability, which is consistent with what was found by (Liu 2004) for Vietnam, it significantly increases SAGE. The positive impact of school travel time might result from the absence of school characteristics in the model or the self-selection of students. For example, a good quality school, which is likely to be located far from home, (i) can facilitate students' learning, and thus improve their SAGE index, and (ii) tend to be chosen by good students, who have high SAGE indexes[9]. The final difference between estimation results for enrolment and SAGE concerns the role of the local environment. In spite of being insignificant in the enrolment decision, variables representing commune characteristics significantly affect schooling progress.

In the estimation for Kinh children only, explanatory variables exhibit a similar pattern of impact to those in the model for all children. For non-Kinh children, only the height-for-age z score, mother's education, household economic condition and distance to school significantly influence their schooling progress.

An important point emerging from comparison between estimations of enrolment and SAGE for non-Kinh children is the difference in the effects of father

9. Although the absence of school characteristics might cause bias in estimated school travel time, this bias is believed to have minor effects on decomposition results, which are the main interest of this study. As shown in the decomposition results, school travel time is not an important contributor to the ethnic gap in schooling progress. Moreover, the estimations of maths and Vietnamese test score equations (Table 4.7 and Table 4.8) show that school travel time is insignificant when school characteristics are controlled for.

and mother's education. Of the two parents' education, only father's years of schooling are significant in the equation of enrolment and only mother's years of schooling can help to increase SAGE. This difference can be attributed to the fact that fathers are more likely to make decisions , e.g. enrolment decisions, in the family, especially in a low income family (Nørlund, Gates, and Vu 1995; Nguyen et al. 2012) like non-Kinh, and mothers are more likely to spend time helping children with their homework or to give them support during the studying process (Nørlund, Gates, and Vu 1995; Nguyen et al. 2012).[10]

4.4.2.2. Decomposition results

The total explained part accounts for 106.5 per cent of the gap in SAGE[11]. Over 100 per cent of the ethnic gap explained in the model implies that unobservable factors help to reduce the gap by 6.5 per cent. The four main factors responsible for the ethnic gap are the height-for-age z score, mother's education, economic condition and commune characteristics. These factors account for 31, 29, 21 and 26 per cent of the gap, respectively. A comparison between decomposition results for enrolment and SAGE shows that the ethnic gap in enrolment is mostly explained by household characteristics, of which father's education and household economic situation are the two dominant factors, while all child, household as well as

10. In the data, 87% and 94% of the household heads are male in Kinh and non-Kinh groups, respectively. The proportion of younger cohort children seeing their mothers daily is 93% compared to 85% of them who see their fathers daily. Most of the children, 96%, have their mother as their primary caregiver.

11. The decomposition for the Tobit model of SAGE shows that the total explained part accounts for 102% of the gap.

commune-related attributes contribute considerably to the gap in schooling progress.

An interesting finding concerns the variation in the contribution of fathers' education and mothers' education to the ethnic gaps in enrolment and SAGE, which results from the variation in the impacts that fathers' education and mothers' education creates in the estimations. In particular, father's education accounts for 51 per cent of the ethnic gap in enrolment but only under 6 per cent in SAGE. In contrast, the explanatory power of mother's education in the latter is more than double that in the former[12].

4.4.3. Schooling performance

The estimation and decomposition results for multilevel models of maths and Vietnamese test scores are presented in Table 4.6 and Table 4.7, respectively.

12. To check whether this finding is driven by a high correlation between mother's education and father's education as suggested by Becker (1973) as regards assortative mating between men and women, education of the father and education of mother is in turn excluded from estimations (see Table A12 and Table A13). The results are still consistent with those when education of both parents are included: father's years of schooling are the dominant factor explaining the enrolment gap and a minor contributor to the SAGE gap; mother's education is insignificant in the enrolment gap but significantly contribute to the SAGE gap.

Table 4.6. Multilevel Models and Decomposition Results for Maths Test Scores

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Minority ethnic child	-1.699***				
	(0.537)				
Age in month	0.057***	0.047***	0.101***	-0.094***	-2.1
	(0.014)	(0.016)	(0.031)	(0.035)	
Boy	-0.309**	-0.326*	0.014	0.005	0.1
	(0.157)	(0.181)	(0.308)	(0.009)	
Health problem	-0.829***	-0.994***	0.452	-0.010	-0.2
	(0.185)	(0.192)	(0.487)	(0.019)	
Grade repetition	-3.112***	-3.059***	-2.915***	0.141***	3.2
	(0.421)	(0.504)	(0.568)	(0.049)	
Speaking Vietnamese at home	1.023**	2.107***	0.046	1.053***	23.6
	(0.427)	(0.687)	(0.383)	(0.231)	
Mother's years of school	0.067**	0.065*	0.023	0.347***	7.8

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.031)	(0.033)	(0.072)	(0.112)	
Unknown mother's education	0.408	0.257	1.317**	0.033	0.7
	(0.351)	(0.404)	(0.669)	(0.026)	
Father's years of school	0.048**	0.056**	0.019	0.145*	3.3
	(0.023)	(0.025)	(0.061)	(0.079)	
Unknown father’s education	0.581**	0.745***	-0.178	0.055*	1.2
	(0.254)	(0.289)	(0.605)	(0.032)	
Older siblings	-0.098	-0.145**	0.064	0.052*	1.2
	(0.062)	(0.068)	(0.122)	(0.030)	
Asset index	0.249	-0.726	7.752**	0.029	0.6
	(2.020)	(2.079)	(3.771)	(0.042)	
School travel time	-0.013	-0.026*	0.002	0.078	1.8
	(0.013)	(0.014)	(0.020)	(0.052)	
Days absent from school of classmates	-0.775	0.104	-5.127***	0.149**	3.3

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.574)	(0.572)	(1.086)	(0.066)	
Grade repetition of classmates	-7.990*	-6.970*	-13.278**	0.273***	6.1
	(4.248)	(3.967)	(6.155)	(0.057)	
Class size	-0.025	-0.054	-0.266**	-0.066	-1.5
	(0.039)	(0.033)	(0.127)	(0.154)	
Television in classroom	2.005**	2.381***	13.993***	0.182***	4.1
	(0.845)	(0.789)	(2.311)	(0.039)	
Dropout at school	-2.363	-2.259*	-3.124	0.384***	8.6
	(1.584)	(1.278)	(3.137)	(0.076)	
School size	0.166***	0.198***	0.300	0.648***	14.5
	(0.056)	(0.044)	(0.248)	(0.082)	
Constant	10.048***	11.547***	3.327		
	(2.563)	(2.824)	(5.150)		
Total				3.404	76.3

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Mean predicted Maths score	14.454	18.883	13.875		
Ln(δ_v)	0.670**	0.333	1.406***		
	(0.293)	(0.317)	(0.147)		
Ln(δ_u)	0.901***	0.837***	-13.172		
	(0.134)	(0.122)	(24.467)		
Ln(δ_ε)	1.390***	1.399***	1.238***		
	(0.028)	(0.029)	(0.068)		
Observations	3,218	2,831	387		

Notes: Data source: 2011-2012 YLSS; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; Standard errors in the decomposition results are bootstrapped with 1000 replications

Table 4.7. Multilevel Models and Decomposition Results for Vietnamese Test Scores

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Minority ethnic child	-1.589***				
	(0.452)				
Age in month	0.054***	0.064***	0.028	-0.089***	-1.9
	(0.018)	(0.022)	(0.020)	(0.032)	
Boy	-1.229***	-1.284***	-0.701**	0.018	0.4
	(0.160)	(0.176)	(0.340)	(0.035)	
Health problem	-0.557***	-0.604***	-0.277	-0.007	-0.1
	(0.181)	(0.181)	(0.575)	(0.013)	
Grade repetition	-2.601***	-2.816***	-1.032*	0.118***	2.5
	(0.445)	(0.533)	(0.571)	(0.042)	
Speaking Vietnamese at home	1.979***	2.351***	1.465**	1.627***	34.8
	(0.462)	(0.637)	(0.664)	(0.237)	
Mother's years of school	0.063***	0.054**	0.023	0.330***	7.1

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.024)	(0.024)	(0.073)	(0.108)	
Unknown mother's education	0.308	0.073	1.104	0.026	0.6
	(0.340)	(0.333)	(1.029)	(0.023)	
Father's years of school	0.084***	0.094***	0.088	0.257***	5.5
	(0.025)	(0.027)	(0.077)	(0.077)	
Unknown father’s education	1.138***	1.392***	-0.275	0.104***	2.2
	(0.311)	(0.319)	(0.738)	(0.036)	
Older siblings	-0.207***	-0.207***	-0.244**	0.103***	2.2
	(0.070)	(0.080)	(0.115)	(0.035)	
Asset index	0.561	0.190	5.443*	0.046	1.0
	(1.653)	(1.729)	(3.081)	(0.040)	
School travel time	0.005	0.003	0.013	-0.011	-0.2
	(0.010)	(0.011)	(0.018)	(0.047)	
Days absent from school of classmates	-0.802	-0.266	-2.566**	0.154**	3.3

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
	(0.504)	(0.589)	(1.115)	(0.065)	
Grade repetition of classmates	-3.083	-1.937	-0.872	0.118**	2.5
	(2.999)	(3.072)	(4.349)	(0.046)	
Class size	-0.010	-0.031	-0.074	0.090	1.9
	(0.030)	(0.030)	(0.078)	(0.150)	
Television in classroom	0.917	1.156	8.327***	0.078**	1.7
	(0.745)	(0.734)	(1.884)	(0.035)	
Dropout at school	-0.345	0.337	-2.192**	0.060	1.3
	(0.980)	(1.140)	(1.061)	(0.052)	
School size	0.139***	0.160***	0.197	0.543***	11.6
	(0.041)	(0.043)	(0.152)	(0.076)	
Constant	10.811***	9.817***	10.841***		
	(2.443)	(2.684)	(3.544)		
Total				3.567	76.4

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	% Explained
Mean predicted Vietnamese scores	20.281	20.897	16.196		
$Ln(\delta_v)$	0.368	0.262	0.981***		
	(0.246)	(0.228)	(0.156)		
$Ln(\delta_u)$	0.551***	0.522***	-15.038		
	(0.102)	(0.096)	(41.481)		
$Ln(\delta_\varepsilon)$	1.363***	1.364***	1.287***		
	(0.019)	(0.020)	(0.053)		
Observations	3,218	2,831	387		

Notes: Data source: 2011-2012 YLSS; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; Standard errors in the decomposition results are bootstrapped with 1000 replications

4.4.3.1. Estimation results

The estimation results for maths and Vietnamese test scores are quite similar. After controlling for other variables, there are significant ethnic gaps in test scores. In particular, both maths and Vietnamese test scores for Kinh children are, on average, over 1.6 point higher than those for non-Kinh children. Factors consistently reducing the scores are male students, health problems, number of siblings and grade repetition of classmates. Variables that help to improve test scores are speaking Vietnamese at home and parents' education[13].

Random effects in the multilevel estimations show the contribution of between-group differences to the total unexplained variance in test scores. In general, class random effects are stronger than school random effects and between-group differences in maths test scores account for a larger share than those of Vietnamese test scores. In particular, 29 and 25 per cent of the total residual variance in the maths test scores is due to between-class and school differences, respectively. For Vietnamese test scores, the comparable figures are 22 and 19 per cent, respectively.

13. The positive sign of the dummy variable for parents' education that is unknown, despite being insignificant in some cases, shows that parents' education unknown by children has a stronger influence on test scores than that known by children. A further investigation shows that Kinh and older children, and children speaking Vietnamese at home, are less likely to know their parents' education (Table A14). Hence, the stronger impact of unknown parents' education might be partly due to a positive association between test scores and Kinh ethnicity, age in month, and speaking Vietnamese at home.

Two remarkable differences between the estimation results for Kinh and non-Kinh children, which are relevant for policy, concern the impacts of the household economic situation and the child characteristics. First, the asset index, despite playing no role in the equation for Kinh children, consistently increases the test scores of non-Kinh children. This finding implies that it is possible to progress the performance of minority children at school by improving their household economic condition, given the fact that ethnic minorities still heavily suffer from poverty. Second, most of the variables related to the child characteristics have strong impacts on test scores of Kinh children but slightly, even insignificantly, affect learning outcomes of non-Kinh children. This suggests that the education system that Kinh children experience allows them to make use of their personality traits in the learning process, resulting in a strong association between children's characteristics and learning outcomes. Non-Kinh children, however, seem to study in a disadvantaged educational environment, in which personal characteristics play a minor role in students' performance. Instead, the test scores of non-Kinh children are likely to be determined by external factors, e.g. household economic condition, peer, class and school effects. Hence, the obstacles to minority children obtaining high test scores are not internal factors, e.g. their characteristics, but disadvantages related to the external factors including the educational environment.

4.4.3.2. Decomposition results

The decomposition results show that the models can explain approximately 73 and 75 per cent of the ethnic gaps in maths and Vietnamese test scores, respectively. There are four groups of variables that mainly contribute to the gaps: use of the Vietnamese language, parents' education, peer effects, and class and school

characteristics. The role of the four groups in explaining the ethnic gaps differs between maths and Vietnamese. For maths scores, the decomposition results show that the two most important factors explaining the ethnic gap are class and school characteristics and speaking Vietnamese at home, individually responsible for around 21 and 25 per cent of the ethnic gap, respectively. The remaining gap is attributable to parents' education, 13 per cent, peer effect, 12 per cent. For Vietnamese test scores, the largest contributor to the ethnic gap is the use of the Vietnamese language, which explains more than one third of the gap. This is followed by parents' education, approximately 16 per cent, school characteristics and peer effects, which together account for 19 per cent of the gap.

Recall that most of the child characteristics are insignificant in the estimation of test scores for the non-Kinh group. Correspondingly, they negligibly explain the ethnic gap in decomposition. Therefore, in order that minority ethnic children catch up with their majority peers at school, the relevant policy should focus on the external factors, e.g. the four groups of contributors that are mostly responsible for the ethnic disparity in education performance.

Finally, there is a noticeable decline in the contribution of the household economic situation and parents' education to the ethnic gaps in test scores compared to that in enrolment and SAGE. Specifically, the asset index explains over a third of the enrolment gap, over a fifth of the SAGE gap and insignificantly contributes to the test score gaps. Similarly, the explanatory power of both parents' years of schooling together is 63, 34 and around 15 per cent in the models of enrolment, SAGE and test scores, respectively. A possible reason for the variation in the contribution of economic condition as well as parents' education is the conceptual

difference between education outcomes under examination. School enrolment, for example, is likely to be related to a household's decision on whether to invest in the child's human capital. Hence, such a decision is considerably determined by characteristics of the main decision maker in the household, e.g. a father's education, and household economic situation, e.g. the asset index. However, whether the child can maintain proper progress at school, represented by SAGE, is affected by other additional factors, such as learning support from the family or a certain effort of the child to pass the exams. Thus, for schooling progress, mother's characteristics are more relevant than father's characteristics, and the asset index becomes less important. When the child has enrolled in school, there is a number of determinants of their performance. For example, apart from the student effort and support from family, peer and school attributes also unduly influence learning outcomes. In this context, the roles of household economic condition and parents' education are further diminished.

4.5. Summary of main findings and policy implications

This paper has investigated primary factors contributing to the education gap between minority and majority ethnic children in Vietnam. The gaps in enrolment, schooling progress and performance were documented and explained through data from the 2009 YLS and 2011-2012 YLSS. The impacts that explanatory variables exert on a child's education vary according to education outcomes and ethnic groups. Some remarkable variations follow. Since fathers and mothers play different roles in the family, especially in minority families, they have strikingly different effects on their child's education. In particular, the father's education has

a positive effect on enrolment of non-Kinh children and only mother's education can help to increase SAGE. Poverty is still an obstacle to minority children obtaining high test scores. Furthermore, for Kinh children, most of the variables related to the child characteristics significantly affect their test scores. For non-Kinh children, however, their performance is likely to be determined by external factors, e.g. household economic condition, home support, peer, class and school effects, rather than their own characteristics. This finding suggests that the factors mainly driving the poor performance of minority children are not their own internal attributes, but disadvantages related to the external determinants.

To identify the key contributors to ethnic inequality, the paper decomposed education gaps between minority and majority ethnic children. The results showed that the key factors contributing to the gaps are different, depending on the type of education outcomes under examination. While ethnic differences in enrolment are entirely explained by variables in the model, around 10 and 20 per cent of the difference in schooling progress and performance are unexplained, respectively. For the enrolment gap, household characteristics are the dominant explanatory factors. However, all child, household and commune attributes are responsible for the gap in schooling progress. Consistent with the insignificance of the child characteristics in the estimation for the non-Kinh group, such characteristics only negligibly explain the test score gaps in decomposition results. Instead, the test score gaps are attributable to the ethnic differences in parent's education, the use of the Vietnamese language, peer and school characteristics.

There are interesting findings relating to the variation in the contribution of explanatory factors to the ethnic gaps. First, father's education is the largest

contributor to the enrolment gap and only plays a minor role in the SAGE gap. In contrast, the role of mother's education is much more important in the latter than in the former. Second, the explanatory power of household economic condition is the largest in the enrolment gap, followed by schooling progress and test scores.

The findings on the variation in the role of determinants suggest efforts to narrow the ethnic gaps in education should vary the focus and priority according to the targeted outcomes. For example, improving household economic condition might be relevant to narrowing the enrolment and schooling progress gaps, but in order to equalize education performance across the ethnic groups, increase Vietnamese language skill among minority ethnic children and improving school quality are more important.

5. Do more educated neighbourhoods experience less property crime? Evidence from Indonesia

5.1. Introduction

The incidence of crime in Indonesia has considerably decreased in recent years. For example, data from Indonesia Social and Economic Surveys (Susenas) show that the share of Indonesians being a victim of crime in the population fell from 3.8 per cent in 2007 to 1.2 per cent in 2012. Reports from the national police headquarters demonstrate that the number of criminals per 100,000 Indonesians declined from 145 to 134 in the same period (BPS 2010, 2013). This decreasing trend in the levels of crime in Indonesia seems to be surprising, given limited capacity and insufficiency of the police force. The number of police per 100,000 Indonesian people is 161 officers, which is just half of the number in the neighbouring country, Thailand (United Nations Office on Drugs and Crime as cited by BMI 2017). And perhaps more importantly, severe corruption among the police force and judiciary is a major obstacle to the fight against criminals. The police force is believed to be the most corrupt among public institutions. According to Transparency International organization, 91 per cent of Indonesians state that police are corrupt or extremely corrupt and the percentage for the judiciary is 86 per cent (TI 2013).

Education levels of Indonesians have consistently improved over time. For instance, gross and net enrolment rates for secondary school over the period 2007-2012 rose by 10 per cent and 7 per cent, respectively, and the proportion of population aged 15 and above with secondary education increased from 27 per cent

to 32 per cent (BPS 2018). Given evidence of the crime reducing impact of education found recently in developed countries such as Sweden (Hjalmarsson, Holmlund, and Lindquist 2015), United Kingdom (Machin, Marie, and Vujić 2011) and the United States (Anderson 2014; Bell, Costa, and Machin 2016; Lochner and Moretti 2004), it is reasonable to ask whether education exerts a similar impact in a developing country such as Indonesia, where law enforcement tends to be weaker and the average education level of the population lower than in developed countries. The more specific question is whether the improvement in education level causally connects to the reduction in the incidence of crime experienced by people in the country during the period 2007-2012.

Recent research establishes that education is able to diminish crime through different channels. For example, education yields an incapacitation effect on crime, especially juvenile crime. An adolescent enrolling in school has less chance of being involved in illegal activities. Supporting evidence of the incapacitation effect was found by Bell, Costa, and Machin (2016). They showed that a longer period of compulsory schooling significantly reduced the crime rate in the United States between 1980 and 2010. Similarly, the expansion of school hours by 30 per cent as a result of a 1997 school reform in Chile led to a decline in all types of youth crimes (Berthelon and Kruger 2011).

There is also suggestive evidence of other channels, in addition to incapacitation, through which education mitigates crime. For example, the change in a conditional cash transfer program in Brazil, which made households with 16-17 year old adolescents enrolling in school eligible for the transfer benefit, helped to lower crime incidents on both school days and non-school days (Chioda, De

Mello, and Soares 2016). Lochner and Moretti (2004) suggest that education is able to decrease crime through preference or income channels. In particular, education increases people's aversion to illegal and risky behaviours, and raises the opportunity cost of crime by creating high potential earnings from lawful activities. Machin, Marie, and Vujić (2012) showed, for instance, that a post-compulsory education policy in the United Kingdom resulted in both a decrease in crime convictions and an increase in the earnings for those affected by the policy. However, the income effect generated by education might also escalate crime by creating high returns to illegal activities (Campaniello, Gray, and Mastrobuoni 2016). Another important channel of impact is the peer effect (Lochner 2011): educated people tend to have well educated friends and colleagues while dropouts are more likely to interact with offenders. In this context, McAdams (2016) found that children who were exposed to older classmates were less likely to commit a crime as adults, and Deming (2011) concluded that a student's own illegal behaviour was significantly associated with her or his delinquent peers.

A key challenge to estimating the effect of education on crime is isolating the causal relationship, which seems to be confounded by unobservable determinants. They include an individual's characteristics such as impatience, returns to crime, ability, attitude toward risk (Lochner and Moretti 2004) or other confounders such as economic shocks and changes in the legal framework. An additional issue associated with identifying the causal relationship concerns the reverse causal effects that unlawful behaviours might impose on educational attainment (Hjalmarsson 2008; Webbink et al. 2012). For instance, juvenile delinquency

significantly reduces the chance of high school and college graduation (Ward and Williams 2015).

To overcome these issues, some studies have successfully exploited variation in education related policies as a source of an exogenous shock to education outcomes. In this regard, the most used educational policy in the literature is that of compulsory schooling. Minimum years required in school, for instance, were used to instrument years of schooling by Lochner and Moretti (2004) and Bell, Costa, and Machin (2016) for the United States, and Hjalmarsson, Holmlund, and Lindquist (2015) for Sweden. Likewise, the minimum school leaving age was made use of by Anderson (2014) and Machin, Marie, and Vujić (2011). The former adopted a triple difference approach to panel data at the county level in the United States, and the latter employed the regression discontinuity method to establish the variation in education outcomes as a result of the change in the minimum school leaving age from 15 to 16 in England and Wales. All aforementioned papers reached conclusive findings related to a negative association between education and crime. Cook and Kang (2016) employed the regulation on school entry cut-off date in the United States, which caused a delay in entering school for children born just after the cut-off date. The author demonstrated that people born just after the cut-off date tended to perform better in school compared to their younger classmates and thus were less likely to commit youth crime. However, these late start students were more likely to commit adult crime due to a higher probability of dropout in the last years of high school. Similarly, McAdams (2016) explained the variation in education outcomes and the probability of incarceration as a result of the regulation on school starting age. The author suggested that raising school starting age would

lower the incarceration rates. Programs related to education have also been exploited to explore the causal impact on crime. For instance, Amin et al. (2016) used Job Corps, a national program providing education and vocational training to disadvantaged youth in the United States, and Deming (2011) employed a public school choice lottery program that gave students the chance of entering their first choice school. Both found that these programs were positively associated with education outcomes and thus mitigated crime.

However, there are several potential gaps in the existing literature. First, an acknowledged limitation of studies using external shocks, for example education related policies, is that the estimated results can only identify the treatment effect locally (Greenstone and Gallagher 2008; Hahn, Todd, and Van der Klaauw 2001). Second, the research so far has addressed the question by using data from police reports or the self reports of criminals, which are likely to be underreported due to the exclusion of unarrested criminals or unreported crime incidents. Finally, most of the previous studies were set in the developed country context[14]. The impact of education on crime, especially property crime, might be different in developing countries, where law enforcement tends to be weaker, poverty higher and the average education level of the population lower than in developed countries.

14. An exception is provided by Berthelon and Kruger (2011) for Chile. Chioda, De Mello, and Soares (2016) also looked at the impact of a conditional cash transfer program on crime in Brazil. However, the result could not be interpreted as the causal effect of education on crime because apart from improving education, the program also increased the incomes of benefited households, which likely, in turn, affected crime. Jonck et al. (2015) found inconclusive results relating to the effect of education on crime in South Africa.

Chapter 5 focuses on examining the causal effect of education on the levels of property crime at the district level in Indonesia. More precisely, the chapter estimates how the education level of people living in a district affects the property crime victimisation rate, defined as the share of people being a victim of property crime in the district population. Hence, information on crime comes from self-reports of individuals who were victims of crimes rather than from police or offender sources. Education, in addition to mitigating criminal activities, also helps to prevent criminal victimisation (Hussin and Zawawi 2012). Highly educated people, for example, probably know good ways to protect themselves from crime and help other neighbours to avoid being victims of crime. However, educated people could earn high incomes and thus tend to be a target of property crime (Carvalho and Lavor 2008). At the district level, living among well-educated neighbours, who are unlikely involved in illegal activities, might reduce the probability of becoming victimised.

The analysis exploits household data from the 2007-2012 Indonesian National Social Economy Survey (Susenas) to create district level panel data. Then, the difference generalised method of moments is adopted to address the endogeneity of education and take into account the dynamics of crime. Property crime, rather than other types of crime, is the focus of this study because over 90 per cent of crime reported in Susenas is property crime. Moreover, conflict related and violent crime, which seem to account for the majority of other crime types[15],

15. Using data from the 2012 Indonesia National Violence Monitoring System (SNPK), I find that conflicts trigger 70% of crime unrelated to violence and 33% of crime related violence. SNPK data were downloaded from

does not solely stem from economic motivations but also from social and political drives. As stated by Kelly (2000), property crime can be plausibly interpreted by economic theory, but violent crime should be examined using social theory, which is beyond the scope of this research.

I find a robust result concerning a negative association between education and property crime in Indonesia. In particular, a one year increase in schooling of the population in a district leads to around a one percentage point reduction in the crime victimisation rate. Primary education appears to play no role in mitigating crime but secondary and higher education does help to alleviate crime. In line with previous research, the results demonstrate a stronger effect of education on crime for males than females. The analyses also provide suggestive evidence of the incapacitation effect for school age children. Further investigation into the variation of the impact shows that extreme poverty weakens the impact of education on crime.

The present paper contributes to the literature in several ways. Both dynamics of crime and endogeneity of education are taken into account by applying difference

http://microdata.worldbank.org/index.php/catalog/2626/get_microdata

GMM[16] to panel data. Also, the present study explores the topic from a different angle, victims of crime rather than criminals, thus contribute to a comprehensive understanding about the relationship between education and crime. Additionally, this paper stands among the few contributions that estimate the effect of education on crime using community level data. It examines the effect of education on crime in a developing country context, which has so far been inadequately studied.

Finally, to the best of my knowledge, this is the first paper identifying the causal relationship between education and crime in Indonesia. A number of previous studies for Indonesia have focused on violence caused by conflicts (Bazzi and Gudgeon 2016; Barron, Kaiser, and Pradhan 2009; Barron and Sharpe 2008; De Juan, Pierskalla, and Vüllers 2015; Mansoob Murshed, Zulfan Tadjoeddin, and Chowdhury 2009; Pierskalla and Sacks 2017; Tadjoeddin and Murshed 2007). Two empirical studies related to crime in the country were those conducted by Cameron and Shah (2013), who investigated how errors in allocating the funds of a poverty program drove crime, and by Hendri and Muharja (2013), who established the

16. Buonanno and Leonida (2006) adopted system GMM to investigate the impact of education on crime in Italy. However, the empirical analysis was conducted by using a small sample of 20 cross sectional observations in 15 periods of time, for which system GMM might yield invalid results. As noted by Blundell and Bond (1998), system GMM is applicable to panel data with short time periods and long cross sections. Moreover, to apply system GMM instead of difference GMM, an additional assumption need to be made: the first differences of endogenous regressors are uncorrelated with the unobservable fixed effect, which seems to be invalid when the time invariant characteristics of districts are likely to affect changes in both education and crime.

impact of poverty and inequality on crime. However, neither of them considered education as an explanatory variable of crime.

5.2. Data

Many of the data used in this paper are taken from the 2007-2012 Indonesia Social and Economic Surveys (Susenas), conducted by Indonesia's Central Bureau of Statistics (BPS) to collect social and economic information about households, including demography, health, education, housing, expenditure and participation in poverty alleviation programs. The data are representative at the district level. The sample size of the 2007-2012 Susenas is presented in Panel A, Table 5

Table 5.1. Descriptive statistics

	2007	2008	2009	2010	2011	2012
A. Sample						
Number of individuals	1,167,019	1,142,675	1,155,566	1,178,493	1,117,827	1,114,445
Number of households	285,186	282,387	291,753	293,715	285,186	286,113
Number of districts	454	456	471	497	497	497
B. Crime						
Crime victimisation rate (%)	3.8	3.12	1.51	1.29	1.24	1.02
Property crime victimisation rate (%)	3.56	2.97	1.41	1.21	1.18	0.93
Share of females in property crime victims	45.91	45.4	37.33	35.56	36.24	37.12
Share of children under 18 in property crime victims	28	26	13	11	14	14
Crime reported rate (%)			16.77	19.45	18.62	17.01
C. Education						
Years of schooling, population aged 7+	7.31	7.33	7.58	7.67	7.68	7.85
Years of schooling, population aged 15+	7.91	7.99	8.22	8.38	8.38	8.58
Years of schooling, females aged 15+	7.41	7.54	7.76	7.96	7.98	8.17
Years of schooling, males aged 15+	8.42	8.45	8.7	8.81	8.79	8.98
Share of population aged 15+ without qualifications	23.01	23.22	22.36	20.01	21.1	19.78

	2007	2008	2009	2010	2011	2012
Share of females aged 15+ without qualifications	27.21	26.67	26.12	23.52	24.57	23.09
Share of males aged 15+ without qualifications	18.68	19.66	18.39	16.44	17.58	16.46
Share of population aged 15+ with primary education	30.43	29.08	29.31	29.72	28.72	28.09
Share of population aged 15+ with junior secondary education	19.83	20.23	19.85	20.57	20.74	21
Share of population aged 15+ with senior secondary education	20.7	21.77	22.09	22.83	22.72	23.97
Share of population aged 15+ with university education	6.03	5.69	6.4	6.87	6.73	7.16
Enrolment rate of population aged 7-18 (%)	84.22	84.41	85	85.56	86.41	87.81
D. Control variables (mean values)						
Population in the districts	496,519	500,515	486,965	478,157	486,014	493,839
Share of males aged 15-25 in districts	19.4	18.86	18.61	18.36	18.32	17.99
Share urban population in districts	34.57	38.23	37.3	37.22	37.26	37.35
Per capita expenditure in districts	297,355	427,403	428,588	473,131	559,154	616,177
Poverty rate in districts	22.77	20.84	19.8	18.47	17.96	17.18
GDP growth rate of districts	5.33	6.61	5.23	5.97	6.36	5.86
Gini index of districts	0.24	0.29	0.28	0.28	0.33	0.34

Data sources: 2007-2012 Susenas and INDO-DAPOER

With regard to information on crime, Susenas includes a question asking if respondents were victims of crime in the last 12 months. Depending on the year of survey, crime is categorised into theft, burglary, murder, fraud, rape, robbery and others. The definition of property crime used in the paper is, however, unchanged over time to include theft, burglary, fraud and robbery. Since 2009, there has been a question asking if the incident was reported to police. Panel B in Table 5.1 shows some descriptive statistics of crime experienced by Indonesians during the period 2007-2012. The shares of the population being a victim of crime (crime victimisation rate) decreased considerably over time. In 2007, 3.8 per cent of the population were victims of crime, which fell to 1 per cent in 2012. More than 90 per cent of the incidents were property crime. Females were less likely to be a target of property crime than males, and the share of female victims tended to decrease over the period. The percentage of property crime victims under 18 year old, despite declining remarkably by half during the period, was still high, at 14 per cent in 2012. Among victims, only less than 20 per cent reported the incidents to police. The low reporting rate partly explains a factor of eight difference between the numbers of self-reported victims in Susenas and crime incidents released by Indonesia's Criminal Police Head Quarters[17]. This also confirms that the volume of crime is, indeed, underreported in police data, thus providing a justification for my use of victim data.

Concerning education, household members aged 5 and above are asked about the highest education level and the corresponding grade that they have completed

17. Data from Criminal Police Head Quarters show that the crime rate, measured by the number of crime incidents per 100.000 Indonesians, was 134 in 2012.

or are currently undertaking. This information enables the construction of a variable identifying the exact years of schooling without measurement errors caused by grade repetition. Panel C, Table 5.1 shows an improvement in the education levels of Indonesians, mostly resulting from a sizable increase in the educational attainment of the younger generation during the period. In particular, the average years of schooling completed by Indonesians rose by 0.54 years over the period. A decline in the Indonesian population without educational qualifications or with primary education coincided with a continuous rise in the share of the population with junior secondary education or above. Comparing education levels between genders, males, on average, had more years of schooling and higher qualifications than females. Among the school age population, the enrolment rates were also increasing regularly over time. Overall, Table 5.1 indicates that the improvement in education was concurrent with the decline in crime during the period. Nevertheless, this only refers to the negative correlation between the two issues. The next sections will identify the causal effect of education on the incidence of crime in the country.

In addition to Susenas data, the empirical analysis also employs data from The Indonesia Database for Policy and Economic Research (INDO-DAPOER), which contains economic and social indicators at the district level.

5.3. Empirical strategy

5.3.1. Model specification

In the empirical analyses, the incidence of crime in a district is measured by the property crime victimisation rate, defined as the number of individuals being a victim of property crime per 100 people[18]. The specific model that estimates the effect of education on crime is as follows.

$$V_{i,t} = \beta_0 + \beta_1 V_{i,t-1} + \beta_2 E_{i,t} + \alpha X_{i,t} + \gamma T_t + \varepsilon_i + u_{i,t} \quad (5.1)$$

Where $V_{i,t}$ and $E_{i,t}$ are the property crime victimisation rate and education level of the population in district *i* in year *t*, respectively; $V_{i,t-1}$ is the property crime victimisation rate in district *i* in year *t-1*; T is a vector of dummy variables controlling for time effects; $X_{i,t}$ is a vector representing the characteristics of district *i* in year *t*; ε_i is unobservable fixed factors and $u_{i,t}$ is unobservable time variant factors. The characteristics of a district include poverty rates and per capita expenditure, which are relevant to economically motivated crime and GDP growth rate and urbanization[19], which represent the development level of the district. The Gini index, calculated from per capita expenditure of households within a district, is added in the model due to recent evidence of correlation between inequality and crime in Indonesia (Cameron and Shah 2013; Hendri and Muharja 2013) as well as other countries (Brush 2007; Kelly 2000). Finally, as suggested by previous

18. BPS (2013) uses the same definition to report crime victims.

19. Data on poverty, GDP and urbanization are taken from Indonesia Database for Policy and Economic Research (DAPOER), http://databank.worldbank.org/data/reports.aspx?source=1266

research using community level data to estimate crime (Anderson 2014; Raphael and Winter-Ebmer 2001), demographic factors such as the size of the population and the share of males aged 15-25 are included in the model. The descriptive statistics of the control variables are presented in Panel D, Table 5.1.

Model (1) is a dynamic panel model, which is relevant to an analysis of crime because the amount of crime in the present is likely to be determined by the levels in the past (Lochner and Moretti 2004). First, from a criminal perspective, undetected illegal activities in the past might encourage offenders to commit further unlawful activities. Second, criminals could accumulate skills and raise returns to crime over time, thus making crime increasingly attractive to them. Third, accumulated skills might also help criminals reduce the probability of being arrested, and as a result, making them persistently commit crimes (Buonanno and Leonida 2006). Finally, from a victim perspective, a person who used to be a victim of crime might learn from the experience to protect themselves better in the future.

5.3.2. Endogeneity

The estimated coefficient of the education variable in equation (5.1) might be biased due to unobservable factors, such as traditional and cultural factors, that affect both the incidence of crime and education level of people in the district. Taking the advantage of the district level panel data, equation (5.1) is differenced to eliminate unobservable fixed factors ε_i.

$$\Delta V_{i,t} = \beta_1 \Delta V_{i,t-1} + \beta_2 \Delta E_{i,t} + \alpha \Delta X_{i,t} + \gamma \Delta T_t + \Delta u_{i,t} \quad (5.2)$$

However, the first difference transformation of a dynamic equation, as in equation (5.2), raises another source of endogeneity. Specifically, on the right hand

side of equation (5.2) $\Delta V_{i,t-1}$ and error terms $\Delta u_{i,t}$ are correlated due to correlation between $V_{i,t-1}$ and $u_{i,t-1}$. That is

$$\text{cov}(\Delta V_{i,t-1}, \Delta u_{i,t}) = \text{cov}(V_{i,t-1} - V_{i,t-2}, u_{i,t} - u_{i,t-1}) \neq 0 .$$

In dynamic panel model (1), $V_{i,t-1}$ is assumed to be a predetermined variable, implying $\text{cov}(V_{i,t-1}, u_{i,t+s}) = 0$ for $s = 0,1,...,T$ s. This assumption can be interpreted as that the current and future disturbances have no effect on the levels of crime in the past. Given unobservable fixed factors being removed, the lags two onward of the crime variable, such as $V_{i,t-2}$, $V_{i,t-3}$, satisfy the relevant and exclusion conditions, and therefore, can be used to instrument $\Delta V_{i,t-1}$. In particular, the lags seem to be (i) highly correlated with the crime variable on the right hand side of equation (5.2), $\text{cov}(V_{i,t-2}, \Delta V_{i,t-1}) = \text{cov}(V_{i,t-2}, V_{i,t-1} - V_{i,t-2}) \neq 0$, and (ii) uncorrelated with the composition of the error terms, $\text{cov}(V_{i,t-2}, \Delta u_{i,t}) = \text{cov}(V_{i,t-2}, u_{i,t} - u_{i,t-1}) = 0$, under the assumption that $V_{i,t-1}$ is a predetermined variable.

Education in equation (5.1) also potentially suffers from endogeneity due to unobservable time variant factors. For example, changes in the legal framework or the enhancement of law enforcement might make criminal behaviour less attractive compared to education for teenage delinquents; or shocks in the labour market, such as an increase in returns to education, also affect relative preference between education and illegal activities. In this case, the first difference transformation of equation (5.1), as shown in equation (5.2), fails to solve the problem of endogeneity because $\Delta E_{i,t}$ and $\Delta u_{i,t}$ on the right hand side of equation (5.2) are correlated. With the same explanation as that for crime, the lags two onward of education seem to

be valid instrumentals for $\Delta E_{i,t}$. In particular, the current and future shocks are unlikely to affect education levels in the past, so $\text{cov}(E_{i,t-2}, \Delta u_{i,t}) = \text{cov}(E_{i,t-2}, u_{i,t} - u_{i,t-1}) = 0$; and the education level of the population as a whole is accumulated over periods of time, so $\text{cov}(E_{i,t-2}, \Delta E_{i,t}) = \text{cov}(E_{i,t-2}, E_{i,t} - E_{i,t-1}) \neq 0$. The instrumental variable estimation of equation (5.2) can be obtained by using the generalised method of moments.

The approach that adopts the generalised method of moments to estimate a dynamic panel model with endogenous explanatory and predetermined variables in differenced form instrumented by their lagged values is known as the difference generalised method of moments (difference GMM) (Arellano and Bond 1991; Roodman 2009). Difference GMM was designed for models with a large number of cross sectional observations and short time periods.

5.3.3. Tests for validity of instrumental variables and robustness of estimation results

Along with the difference GMM estimators of crime, various tests are conducted to verify the validity of instrumental variables (Roodman 2009). The Sargan and Hansen tests of overidentifying restrictions are used to test the null hypothesis that instrumental variables are jointly exogenous. Another test concerning the exogenous condition of instrumental variables, which especially applies to difference GMM estimations, is that of no serial correlation of order one between error terms in the original equation, equation (5.1). If the null hypothesis is not

rejected, then the second lag of the crime victimisation rates, $V_{i,t-2}$, can be used to instrument its first lag in differenced form, $\Delta V_{i,t-1}$, in equation (5.2)[20].

Roodman (2009) acknowledged the potential bias of difference GMM in finite samples, especially when the number of instrumental variables is large. According to Roodman (2009), a large set of instruments is likely to weaken the Hansen test of overidentifying restrictions as well as overfit endogenous explanatory variables. To address this issues, the number of instrumental variables is reduced by using the collapsed lags two to four of the crime and education variables to instrument their differenced form. This set of instrumental variables is also employed as the main specification in the empirical analysis. In addition, various combinations of lags are used to check the sensitivity of estimation results.

5.4. Empirical results

5.4.1. Benchmark estimations

This section discusses the full estimation results by interpreting the effects of education and other covariates on property crime victimisation rates. Collapsed lags two to four of the variables in level are used to instrument the endogenous variables in the differenced form and Windmeijer (2005) finite sample correction is adopted

20. Since equation (5.2) is the first difference transformation of equation (5.1), no serial correlation of order 1 between error terms in (1) implies no serial correlation of order two in equation (5.2). Alternatively, $\text{cov}(u_{i,t}, u_{i,t-1}) = 0$ in (4.1) is equivalent to $\text{cov}(\Delta u_{i,t}, \Delta u_{i,t-2}) = \text{cov}(u_{i,t} - u_{i,t-1}, u_{i,t-2} - u_{i,t-3}) = 0$ in (4.2).

to generate robust standard errors. The education variable in the benchmark estimations and in most other estimations is years of schooling of the population aged 15 and above. Recall that years of schooling derived from Susenas do not include grade repetition, so they indicate the exact educational attainment of the population in year. The population aged 15 and above is the focus of attention because people in this age group are more likely to be involved in crime incidents. Nevertheless, the estimation results using other indicators of education and for different groups of population are presented in next sections. The point estimates and associated standard errors of education and other control variables are presented in Table 5.2. Only time effect is controlled for in column 1; demographic factors are included in column 2; economic variables are added in column 3; and column 4 displays the full model.

Table 5.2. Benchmark estimations

Victim rate	(1)	(2)	(3)	(4)
Lagged victim rate	0.117***	0.122***	0.110***	0.113***
	(0.033)	(0.033)	(0.033)	(0.032)
Years of schooling, aged 15+	-0.854**	-0.853***	-0.929**	-0.932***
	(0.335)	(0.330)	(0.365)	(0.352)
Year = 2009	-1.326***	-1.344***	-1.310***	-1.329***
	(0.116)	(0.115)	(0.117)	(0.115)
Year = 2010	-0.932***	-0.979***	-0.989***	-1.045***
	(0.226)	(0.216)	(0.224)	(0.210)
Year = 2011	-1.177***	-1.245***	-1.461***	-1.549***
	(0.167)	(0.158)	(0.155)	(0.151)
Year = 2012	-1.244***	-1.330***	-1.606***	-1.714***
	(0.216)	(0.203)	(0.195)	(0.184)
Ln(population)		0.944***		1.127***

Victim rate	(1)	(2)	(3)	(4)
		(0.360)		(0.371)
Urban population (%)		0.005		0.004
		(0.005)		(0.005)
Male, aged 15-25 (%)		-0.009		-0.010
		(0.019)		(0.019)
Ln(per capita expenditure)			0.770**	0.788**
			(0.366)	(0.368)
Gini			1.791**	1.854**
			(0.873)	(0.869)
Poverty (%)			0.001	0.001
			(0.002)	(0.002)
GDP growth (%)			0.002	0.002
			(0.005)	(0.005)
Number of instruments[21]	10	13	14	17
Sargan test (p value)	0.655	0.563	0.723	0.607
Hansen test (p value)	0.542	0.447	0.728	0.602
Test of AR(1) (p value)	0.848	0.774	0.845	0.782
Number of districts[22]	507	507	507	507
Number of observations	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A15.

The results are consistent across specifications, showing that more educated districts experience less property crime. On average, a one year increase in

[21] Numbers of instrumental variables reported in diff-GMM include numbers of exogenous explanatory variables served as instrumental variables. Hence, the lagged endogenous explanatory variables served as instrumental variables are fewer than the reported numbers, e.g 6 in the full model (4).

[22] During period 2007-2012, there was district proliferation, so numbers of districts were different between years (Table 5.1). Numbers of districts in diff-GMM estimations, e.g. Table 5.2, are the total numbers of districts that ever existed in the period.

schooling of the population leads to almost a 0.93 percentage point decrease in the victim rate. Given the fact that the average years of schooling of Indonesians rise by 0.5 years over the six year period, from 2007 to 2012, and the average share of Indonesians being a victim of crime in this period is 1.9 per cent, the magnitude of the crime reducing impact that education creates in Indonesia is sizable. It can be said that the improvement in education levels of Indonesians significantly contribute to the reduction in the incidence of property crime in the country.

As shown in all estimations, the incidence of crime in a year significantly depends on the amount of crime occurring in the previous year. The negative coefficients of the dummy variables controlling for time effects indicate a decline in the property crime victimisation rates during the period. Among variables characterizing a district's development, the Gini index and average per capital expenditure are positively associated with the levels of crime, consistent with Cameron and Shah (2013). Other variables including poverty rates, GDP growth rates, urbanization and the share of males aged 15-25 in the population have no significant influence on the crime situation in the district, but districts with large population experience more property crimes relative to small districts[23]. Finally, in all four specifications, the Sargan and Hansen tests of overidentifying restrictions

[23] There is evidence about the association between unemployment and crime (Melick 2003; Papps and Winkelmann 2000). However, information on unemployment at the district level in Indonesia is not accessible. Thus, in Table A16 in Appendix I use the share of inactive population in a district to represent unemployment. The results show that the share of inactive population is insignificant.

and the test of no serial correlation of order one between error terms speak to the statistical validity of the instrumental variables.

5.4.2. Robustness checks[24]

This section checks the sensitivity of difference GMM estimators in dynamic panel models to the length of lags. Various estimations are conducted to check whether the estimated effects of education on crime are robust (Table 5.3). First, the length of lag is varied so that the instrumental sets include only lag two (column 1), lag two and three (column 2), lag two to four (column 3), and lag two to five (column 4). Then, control variables related to the economic condition of the districts including per capita expenditure, the Gini indexes, poverty rates and GDP growth rates are treated as endogenous regressors (column 5), rather than as exogenous. Since the number of endogenous explanatory variables considerably rises in this specification (column 5), the combination of collapsed lags 2-4 is made use of to reduce the number of instrumental variables[25]. All estimations in this section employ the full set of control variables, as per that in column 4, Table 5.2.

24. I replicate all of these estimations for the restricted sample where districts in Jakarta are excluded. Jakarta is often excluded in empirical analyses using the district level data, for example Fitrani, Hofman, and Kaiser (2005) and Lewis (2017), because its districts are considered outliners. The results are not sensitive to the exclusion of Jakarta.

25. I treat economic-related variables as exogenous regressors in most of the specifications because this help to reduce the number of instrumental variables, thus substantially diminishing bias Roodman (2009).

Table 5.3. Robustness check

Victimisation rate	(1) Lag 2 only	(2) Lag 2-3	(3) Lag 2-4	(4) Lag 2-5	(5) Collapsed lags 2-4
Years of schooling, aged 15+	-0.866***	-0.909***	-0.932***	-0.866***	-0.929***
	(0.200)	(0.198)	(0.200)	(0.196)	(0.253)
Number of instruments	19	25	29	31	25
Sargan test (p value)	0.527	0.225	0.380	0.441	0.868
Hansen test (p value)	0.542	0.221	0.437	0.198	0.714
Test of AR(1) (p value)	0.817	0.839	0.854	0.786	0.684
Number of observations	1,812	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A17.

The results consistently show the significant effects of education on the amount of crime in a district. Moreover, the magnitude of the effects is stable across specifications, ranging from 0.866 to 0.932, which is similar to that in the benchmark models. The Sargan and Hansen tests and the test of serial correlation imply the validity of the instrumental variables.

In short, the point estimates of education are not sensitive to the length of lags served as instrument variables in difference GMM, confirming the robust effects of education on crime reduction in Indonesia.

5.4.3. Different indicators of education

This section adopts the same specification to that in column 4, Table 5.2 (the full set of control variables and lags two to two serving as instrumental variables) but uses different indicators of education to further explore the crime reducing potential of education (see Table 5.2). First, years of schooling of the population aged seven and above, rather than those of age group 15 and above, are used to explain the variation in the victim rate (column 1). Still, the effect is highly significant, indicating that if all Indonesians extend the time they stay in school by one year, the property crime victimisation rate will decrease by 1.1 percentage point.

Next, the share of the population aged 15 and above without educational qualifications is employed to verify whether this variable is positively associated with the volume of crime, as found by Machin, Marie, and Vujić (2011) for England and Wales. The estimation results show that an increase of one percentage point in the share of district population without qualifications indeed leads to additional 0.13 percentage points of the property crime victimisation rates (column 2).

Table 5.4. Different measurements of education

Victimisation rate	(1)	(2)	(3)
Years of schooling, aged 7+	-1.074***		
	(0.417)		
No qualifications, aged 15+ (%)		0.122***	
		(0.040)	
Primary education, aged 15+ (%)			0.066
			(0.082)
Junior Secondary education, aged 15+ (%)			-0.134*
			(0.071)
Senior Secondary education and above, aged 15+ (%)			-0.172***
			(0.062)
Number of instruments	17	17	23
Sargan test (p value)	0.522	0.378	0.617
Hansen test (p value)	0.531	0.128	0.367
Test of AR(1) (p value)	0.612	0.856	0.918
Number of observations	1,812	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A18.

Finally, no educational qualifications is served as the reference category for comparison with other levels of education (column 3). The results suggest that the impacts of education on crime rise with levels of education. Specifically, the change in the share of the population aged 15 and above attaining primary education has no effect on the incidence of property crime occurring in the district. However, a one percentage point increase in the population obtaining junior and senior secondary education leads to a decrease of 0.13 and almost 0.17 percentage point in the victim rates, respectively. It appears that completing 6 years of primary

education does not generate a significant non-pecuniary benefit relating crime reduction for Indonesia, but junior secondary and higher education does.

In summary, the causal effect of education on crime is found robust across different indicators of education and senior secondary education appears to play a particular important role in mitigating property crime in Indonesia.

5.4.4. School age children

As mentioned earlier, the shares of children under 18 year old among the victims of property crime in Indonesia declined remarkably by half during the period 2007-2012 while the enrolment rates among school age children gradually rose from 84.2 per cent to 87.8 per cent (Table 5.1). Thus, this section examines whether enrolling in a school can help children to avoid being victimised and whether children's enrolment can also lessen the amount of crime experienced by people in the district.

Table 5.5. School age children

Victimisation rate	(1) All population	(2) Aged 7-18
Enrolment rate, aged 7-18	-0.088**	0.012
	(0.044)	(0.041)
Number of instruments	17	17
Sargan test (p value)	0.338	0.513
Hansen test (p value)	0.108	0.220
Test of AR(1) (p value)	0.187	0.257
Number of observations	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A19.

Table 5.5 presents the estimation results when the school enrolment rates among the school age population are used to represent the level of education in the district. In column 1, the crime variable is the property crime victimisation rate of the whole population in the district. The result shows that districts with high school enrolment rates experience less property crime than those with low enrolment rates. An additional one percentage point of the population aged 7-18 staying in school causes a 0.09 percentage point decrease in the victim rate. A possible explanation is that students are supervised in school, so they are less likely to be victims of crime, as a result, decreasing the victim rate in the district as a whole. An alternative explanation concerns the incapacitation impact of schooling on crime. Students are unable to present in classes and participate in illegal activities at the same time. The estimation result in column 2 suggests that the latter interpretation is likely to be reasonable in this case. In column 2 where the levels of crime are measured by the property crime victimisation rate of the school age population, the coefficient of enrolment is insignificant. Hence, enrolling in school does not help to protect the school age cohort from being a victim of property crime, implying that the former interpretation does not fit.

In short, the results suggest children's school enrolment, despite not significantly lessening the likelihood of being victimised for children themselves, yields an incapacitation impact, through which it prevents children from participating in illegal activities, and thus alleviates the incidence of crime.

5.4.5. Gender difference

Table 5.6. Gender difference

Victimisation rate	(1) All population	(2) Females	(3) All population	(4) Males
Years of schooling, female aged 15+	**-0.834*** (0.314)**	**-0.499* (0.299)**		
Years of schooling, male aged 15+			**-1.011** (0.413)**	**-1.383** (0.547)**
Number of instruments	17	17	17	17
Sargan test (p value)	0.667	0.401	0.604	0.749
Hansen test (p value)	0.725	0.435	0.508	0.821
Test of AR(1) (p value)	0.885	0.525	0.612	0.879
Observations	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A20.

This section explores how the impacts of education on crime vary by gender. First, the victim rates and education variables are reconstructed separately for females and males in the district. Next, estimations are conducted to gauge how the education levels of females and males affect their own experience of property crime as well as the amount of crime experienced by people in the district as a whole. Recall that Indonesian females, on average, are less likely to be victimised and have lower education than males (Table 5.1). Estimation results by gender are displayed in Table 5.6. Although the education levels of both genders significantly decrease the incidence of crime occurring in the district, the effect exerted by the education of males (column 3) is considerably stronger than that of females (column 1). Similarly, the crime reducing impact of education in the equation for man (column

4) approximately doubles that for women (column 2). This result is in line with the findings concerning the gender difference in arrested rates by Amin et al. (2016) and Anderson (2014) for the United States.

5.4.6. Poverty and the impact of education on property crime

As examining the impact of education on crime, particularly property crime, in a developing country like Indonesia, where over 10 per cent of the population is poor, it is relevant to consider this impact in the connection with poverty. Although no evidence to suggest a significant effect of poverty on crime has been found in the benchmark estimations (Table 5.2), this section gauges whether the effect of education on crime varies relative to the extent of poverty situation in a district. The interactions between the average years of schooling and the share of the population in the district living under a particular cut off expenditure are included in the estimation, as shown in equation (5.3) below.

$$V_{i,t} = \beta_0 + \beta_1 V_{i,t-1} + \beta_2 E_{i,t} + \beta_3 E_{i,t} P_{i,t} + \alpha X_{i,t} + \gamma T_t + \varepsilon_i + u_{i,t} \quad (5.3)$$

Where $P_{i,t}$ is the share of the population in the district living under a particular level of expenditure, for instance the 10th percentile of the country's distribution. In equation (5.3), the total marginal effect of education on crime is $\beta_2 + \beta_3 P_{i,t}$. Hence, the coefficient of the interaction term, β_3, refers to variations in the effect of education on crime relative to the poverty situation in the district.

Table 5.7 summarises the main estimation results of equation (5.3) with different levels of expenditure, ranging from the 5th (column 1) to 20th percentile (column 5). In column 3 is the interaction between education and poverty rates,

which varied between 10 per cent and 15 per cent in 2007-2012. The positive coefficients of the interaction terms between education and the share of the 5 per cent and 10 per cent poorest population living in the district, although only significant at the level of 0.1, implying that poverty weakens the extent to which education affects crime. However, the interactions with the higher cut off levels of expenditure, such as the poverty line, the 15^{th} or 20^{th} percentile, are statistically insignificant. Hence, the results suggest that poverty only weakens the extent to which education alleviate crime in districts with extremely poor households, those living below the 10^{th} percentile of Indonesian expenditure distribution.

Anderson (2014) is among few who have investigated the variation in the effect of education on crime across income levels. However, the author did not find any statistically significant variation. The possible reason is that Anderson (2014) analysed the relationship between education and crime in the context of a developed country like the United States, where the social security system mostly ensures the basic necessities of poor people, thus could not find a difference in the effect of education on crime across income levels.

Table 5.7. Interaction with poverty

	(1)	(2)	(3)	(4)	(5)
Victimisation rate	5th percentile	10th percentile	Poverty line	15th percentile	20th percentile
Years of schooling, aged 15+	-1.247***	-1.114***	-0.988***	-1.027***	-1.183***
	(0.452)	(0.384)	(0.369)	(0.357)	(0.425)
Interaction	0.015*	0.007*	0.002	0.004	0.008
	(0.008)	(0.004)	(0.002)	(0.003)	(0.007)
Poverty rate	0.002	0.002	-0.019	0.002	0.002
	(0.002)	(0.002)	(0.018)	(0.002)	(0.002)
Number of instruments	20	20	20	20	20
Sargan test (p value)	0.978	0.989	0.870	0.991	0.911
Hansen test (p value)	0.932	0.944	0.744	0.904	0.634
Test of AR(1) (p value)	0.615	0.921	0.799	0.783	0.824
Observations	1,812	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas and INDO-DAPOER; Significance level: *10%; **5%; ***1%; Standard errors in parentheses; The full estimation and test results are presented in Appendix, Table A21.

5.5. Summary of main findings and policy implications

The shares of Indonesians being a victim of crime have considerably declined in recent years, from 3.8 per cent in 2007 to 1.2 per cent in 2012 while the education levels of people in the country have consistently increased over time. This paper established the causal relationship between education and property crime in Indonesia over the period 2007-2012. The victimization rate at the district level was used to represent the incidence of crime. The difference GMM estimation technique was employed to address the endogeneity of education and the dynamics of crime. The results show that districts with a higher average level of education experience less property crime, and this causal relationship is robust across different model specifications and education indicators. In conclusion, the improvement in the education level of Indonesians significantly contributes to the reduction in the incidence of crime in the country.

A further investigation demonstrates that the crime reducing effect generated by males' education is stronger than that generated by females' education. Estimations of crime victimization rates on different indicators of education show that the degree to which education diminishes crime rises with levels of education. While primary education appears to have no influence on the levels of crime, higher education, especially senior secondary education and above, leads to significantly less crime. In addition, there is suggestive evidence that schooling yields an incapacitation impact, through which it prevents children from participating in illegal activities. Finally, extreme poverty can mitigate the effect of education on crime.

The findings relating to the crime reducing impact of education, especially in terms of senior secondary education, in this paper provide an additional supporting reason for the implementation of compulsory senior secondary education in Indonesia, which commenced recently. A key recommendation of this paper would be that the non-pecuniary benefits of schooling such as crime reduction, should be taken into account in the policy making process, for example, policies related to compulsory senior secondary education. Additionally, the government should consider educational improvement and poverty mitigation as possible solutions, apart from law enforcement, to crime.

6. Conclusions

Human capital yields both economic and non-economic benefits, and thus plays an important role in the development of a society. It is even a critical factor for developing countries in their journey to catch up with the developed world. With the objective of obtaining better a understanding of human capital in general, and that in developing countries in particular, this thesis has investigated three different issues related to human capital by using data from Indonesia and Vietnam.

The first theme concerns a determinant of human capital: time allocation. Because human capital is accumulated gradually through learning and working process, the way children spend their time on studying, working and leisure directly affects their physical and mental health, their skills and knowledge, thus determines their human capital. In this regard, Chapter 3 examined how the child's time is allocated between activities. It focused particularly on the role of parental academic aspirations in determining the child's time allocation.

The second and third themes are related to a narrow definition of human capital: education. In particular, Chapter 4 spotlighted inequality in education in Vietnam. This Chapter sought explanations for the ethnic heterogeneity of education outcomes in the country, where the non-Kinh minority children lag far behind in all education indicators compared the Kinh majority peers. Chapter 5 examined a non-economic benefit of education: crime reducing effect. For this theme, Indonesia is an interesting case because this country has experienced a consistent decrease in crime victimization rates, despite an extremely high level of corruption in the police force and justice system. Hence, it is reasonable to question

if an improvement in the education level of Indonesians is a driving factor behind the decrease in the incidence of crime. The main results and findings of each research paper are summarized below.

6.1. Summary

Chapter 3 examined the link between parental academic aspirations for a child and the allocation of a child's time. First, a general answer was derived from the economic model of household utility maximization. In the model, parental aspirations are a factor that augments the efficiency of a child's studying time, which, in turn, helps to increase the human capital that a child accumulates, and thus raise household utility. The statistical analysis shows that parental aspirations are positively associated with the optimal studying time of the child and negatively associated with the optimal working time. To verify the findings from the theoretical model, data from the 2006 and 2014 Vietnam Young Lives Survey were used to estimate a system of equations of a child's time allocation, in which parental aspirations are the key explanatory variable of interest. The results of the empirical analysis support the theoretical findings, confirming a causal relationship between parental academic aspirations and the allocation of a child's time. A further analysis suggests that parental academic aspirations have no significant effect on the time doing housework for boys, but significantly reduce the workload for girls. Academic performance of a child is positively affected by hours of study, and negatively influenced by the time he or she works.

Chapter 4 explained the existence of the disparity in education outcomes between non-Kinh ethnic minority children and the Kinh ethnic majority

counterpart in Vietnam. The ethic gaps in favour of the Kinh group were found for various education indicators including enrolment, schooling progress and academic performance. The investigation used data from the 2009 YLS and 2011-2012 YLSS, and employed Probit and multilevel regression models, and associated decomposition techniques. The results suggest that the key determinants of education outcomes as well as the main factors contributing to the ethnic gaps vary, depending on the education indicators and ethnic groups being examined. Education of the father, who is likely to be the decision maker in the family, significantly affects the enrolment status of the child, while education of the mother, who spends more time with the child, is more important to the child's schooling progress. Disadvantages related to the external determinants, such as poverty, low level of home support and low school quality, are still obstacles to minority children obtaining high test scores. Father's education and household economic condition are dominant contributors to the difference in enrolment rates between Kinh and non-Kinh children. However, the gap in schooling progress is explained by all child, household and commune attributes. The test score gaps are attributable to a broader set of variables including, but not limited to, parent's education, the use of the Vietnamese language, peer and school characteristics.

Chapter 5 examined the effect of education on property crime in Indonesia over the period 2007-2012. The incidence of crime is represented by the victimization rate at the district level. The diff-GMM estimation technique was employed to address the endogeneity of education and dynamics of crime, and thus identifying the causal effect. The results show that more educated districts experience less property crime. In particular, a 1 year rise in schooling of people living in the district, on average, reduces the victimization rate by 1 percentage

point. A further investigation demonstrated that the causal effect is found stronger in the models associated with the level education of the male population and crime incidents reported by males than those associated with females. Secondary and higher education are particularly important to mitigating crime. In addition, there is suggestive evidence that schooling yields an incapacitation effect to prevent children from participating in illegal activities. Finally, poverty appears to have no direct influence on the incidence of crime, but extreme poverty can mitigate the extent to which education reduces crime.

6.2. Contributions and policy implications

The theoretical and empirical analyses conducted in this thesis have made significant contributions to literature about human capital in various aspects. For example, one of the themes in this thesis appears to have never been explored before, i.e. the relationship between parental academic aspirations and a child's time allocation; the determinants of the ethnic gap in education have been precisely identified by investigating different education outcomes; an old research question has been approached from a different aspect, i.e. examining the impacts of education on crime from victims' perspectives rather than criminals. The findings and policy implications related to education and human capital in this thesis are particularly important for Indonesia and Vietnam to develop their economies and societies. The detail contributions and policy implications derived from each research theme are as follows.

Chapter 3's contribution to the literature is certain because it is the first to investigate the relationship between parental academic aspirations for a child and

the child's time allocation. It identified that the allocation of a child's time is a missing link in the relationship between parental aspirations and the child's academic achievements. The findings suggest that future studies should consider the allocation of a child's time as a channel through which parental aspirations affect a child's educational attainment. The findings also have a policy implication that interventions or programs, such as those related to economic and social returns to education, can raise parents' academic expectations and aspirations for a child, and thus positively affect the child outcomes.

Chapter 4 adds to the literature by investigating the ethnic gaps across different education outcomes. The results show that key factors explaining the ethnic gaps in education might be different, depending on the education outcome under examination. These results, despite being straightforwardly relevant for policy implication, have not been explicitly specified in previous studies for Vietnam. The findings suggest that efforts to narrow the ethnic gaps in education should vary the focus and priority according to the targeted outcomes. For example, improving household economic condition might be relevant to narrowing the enrolment and schooling progress gaps, but in order to equalize education performance across the ethnic groups, removing the language barriers and improving school quality are more important.

Chapter 5 contributes to the existing literature by examining the impact of education on crime from a different angle: victims' perspective rather than criminals. The findings relating to the crime reducing impact of education in Chapter 5 constitute strong evidence of non-pecuniary benefits of education. In this regard, the key policy recommendation is that non-pecuniary benefits of schooling

such as crime reduction should be taken into account in the decisions related to education policy. In the case of Indonesia, the findings are supportive evidence for the expansion of basic education through senior secondary school, which has been recently launched. Additionally, in the fight against crime, the government should consider education improvement and poverty mitigation as efficient tools beside law enforcement.

6.3. Limitations of the thesis and suggestions for further studies

The author acknowledges numbers of limitations of the thesis and propose suggestions for further studies as follows.

In chapter 3, the theoretical model was simplified by assuming that leisure time is constant and not distinguishing between market and home productions. These assumptions may not be realistic, but they facilitate the algebraic derivative of the association between parental aspirations and a child's time allocation on studying and working, which is the main purpose of the theoretical analysis. In addition, there may be a criticism about exogeneity of the instrumental variable used in the empirical models, the average local wage. Indeed, the instrumental variable might be not exogenous to the extent that it may be correlated with the economic development level and labor market in a region, which, at the same time, affect a child's work and study. Although controlling for regional dummy variables in the empirical analysis significantly help to reduce bias due to unobservable regional fixed effects, this cannot capture all characteristics of an economic region.

As mentioned, research in chapter 3 is the first to explore the impact of parental academic aspirations on a child's time allocation. This suggests that further comparative work, in other countries, especially in the developing world, would be useful.

In chapter 4, the data used in the main analysis are from the Young Lives surveys. Given the Young Lives sample is not nationally representative, the findings in this chapter should be interpreted with caution. The limitation of the sample suggests that further studies along this line using national representative data would be useful.

In chapter 5, one relevant explanatory variable, unemployment rates, was not included in the econometric models because the data were not accessible at the district level. Including unemployment rates in the empirical models would possibly provide insightful analyses and policy implications. However, the absence of this variable by no means undermines the conclusions related to the impacts of education on crime, which is the main focus of the chapter. Furthermore, the diff-GMM estimation technique employed in the econometric models helps to address potential bias caused by the omission of unemployment rates.

Reference

Acemoglu, Daron, Gallego Francisco A, and Robinson James A. 2014. "Institutions, Human Capital, and Development." *Annu. Rev. Econ.* 6 (1):875-912.

Alcaraz, Carlo, Chiquiar Daniel, and Salcedo Alejandrina. 2012. "Remittances, Schooling, and Child Labor in Mexico." *Journal of development economics* 97 (1):156-65.

Altepeter, Thomas S. 1989. "The Ppvt -R as a Mea Functioning: A Caution." *Journal of Clinical Psychology* 45 (6):935-41.

Amin, Vikesh, Flores Carlos A, Flores-Lagunes Alfonso, and Parisian Daniel J. 2016. "The Effect of Degree Attainment on Arrests: Evidence from a Randomized Social Experiment." *Economics of Education Review* 54:259-73.

Anderson, D Mark. 2014. "In School and out of Trouble? The Minimum Dropout Age and Juvenile Crime." *Review of Economics and Statistics* 96 (2):318-31.

Anderson, Darrell E, and Flax Morton L. 1968. "A Comparison of the Peabody Picture Vocabulary Test with the Wechsler Intelligence Scale for Children." *The Journal of Educational Research* 62 (3):114-6.

Areepattamannil, Shaljan, and Lee Daphnee HL. 2014. "Linking Immigrant Parents' Educational Expectations and Aspirations to Their Children's School Performance." *The Journal of genetic psychology* 175 (1):51-7.

Arellano, Manuel, and Bond Stephen. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The review of economic studies* 58 (2):277-97.

Arouri, Mohamed El Hedi, Ben Yousse Adel, and Nguyen Cuong. 2016. *Ethnic and Racial Disparities in Children's Education: Comparative Evidence from Ethiopia, India, Peru, and Viet Nam*. Helsinki: UNU-WIDER.

Bacolod, Marigee P, and Ranjan Priya. 2008. "Why Children Work, Attend School, or Stay Idle: The Roles of Ability and Household Wealth." *Economic Development and Cultural Change* 56 (4):791-828.

Badiani, Reena, Baulch B, Brandt L, Dat VH, Giang NT, Gibson J, Giles J, Hinsdale I, Hung P, and Kozel V. 2013. *2012 Vietnam Poverty Assessment—Well Begun, Not yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges*. Washington, DC: World Bank.

Baert, Stijn, and Cockx Bart. 2013. "Pure Ethnic Gaps in Educational Attainment and School to Work Transitions: When Do They Arise?" *Economics of Education Review* 36:276-94.

Baker, Frank B. 2001. *The Basics of Item Response Theory*. United States of America ERIC.

Baland, Jean-Marie, and Robinson James A. 2000. "Is Child Labor Inefficient?" American *journal of Political Economy* 108 (4):663-79.

Barakat, Bilal. 2016. "“Sorry I Forgot Your Birthday!”: Adjusting Apparent School Participation for Survey Timing When Age Is Measured in Whole Years." *International Journal of Educational Development* 49:300-13.

Barron, Patrick, Kaiser Kai, and Pradhan Menno. 2009. "Understanding Variations in Local Conflict: Evidence and Implications from Indonesia." *World Development* 37 (3):698-713.

Barron, Patrick, and Sharpe Joanne. 2008. "Local Conflict in Post-Suharto Indonesia: Understanding Variations in Violence Levels and Forms through Local Newspapers." *Journal of East Asian Studies* 8 (3):395-423.

Basu, Kaushik, and Van Pham Hoang. 1998. "The Economics of Child Labor." *American economic review* 88 (3):412-27.

Bauer, Thomas K, and Sinning Mathias. 2010. "Blinder–Oaxaca Decomposition for Tobit Models." *Applied Economics* 42 (12):1569-75.

Baulch, Bob, Hoa Thi Minh Nguyen, Phuong Thi Thu Phuong, and Hung Thai Pham. 2010. *Ethnic Minority Poverty in Vietnam*, *Working Paper No. 169*. Manchester: Chronic Poverty Research Centre

Bazzi, Samuel, and Gudgeon Matthew. 2016. "Local Government Proliferation, Diversity, and Conflict." In.: Boston University-Department of Economics.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political economy* 81 (4):813-46.

Bell, Brian, Costa Rui, and Machin Stephen. 2016. "Crime, Compulsory Schooling Laws and Education." *Economics of Education Review* 54:214-26.

Benhabib, Jess, and Spiegel Mark M. 1994. "The Role of Human Capital in Economic Development Evidence from Aggregate Cross-Country Data." *Journal of Monetary economics* 34 (2):143-73.

Berthelon, Matias E, and Kruger Diana I. 2011. "Risky Behavior among Youth: Incapacitation Effects of School on Adolescent Motherhood and Crime in Chile." *Journal of public economics* 95 (1):41-53.

Bhalotra, Sonia, and Heady Christopher. 2003. "Child Farm Labor: The Wealth Paradox." *The World Bank Economic Review* 17 (2):197-227.

Black, Sandra E, Devereux Paul J, and Salvanes Kjell G. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *The Review of Economics and Statistics* 93 (2):455-67.

Blundell, Richard, and Bond Stephen. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of econometrics* 87 (1):115-43.

BMI. 2017. "Indonesia Crime and Security Risk Report." In. London, United Kingdom: Business Monitor International Ltd.

Bollen, Kenneth A, Glanville Jennifer L, and Stecklov Guy. 2002. "Economic Status Proxies in Studies of Fertility in Developing Countries: Does the Measure Matter?" *Population studies* 56 (1):81-96.

Bouoiyour, Jamal, and Miftah Amal. 2014. "Household Welfare, International Migration and Children Time Allocation in Rural Morocco." *Journal of Economic Development* 39 (2):75-95.

BPS. 2010. "Criminal Statistics 2010." In. Jakarta, Indonesia: The Central Bureau of Statistics.

BPS. 2013. "Criminal Statistics 2013." In. Jakarta, Indonesia: The Central Bureau Of Statistics.

BPS. "Education Indicators, 1994-2016." https://www.bps.go.id/statictable/2010/03/19/1525/indikator-pendidikan-1994-2017.html.

Brush, Jesse. 2007. "Does Income Inequality Lead to More Crime? A Comparison of Cross-Sectional and Time-Series Analyses of United States Counties." *Economics letters* 96 (2):264-8.

Buonanno, Paolo, and Leonida Leone. 2006. "Education and Crime: Evidence from Italian Regions." *Applied Economics Letters* 13 (11):709-13.

Cameron, Lisa, and Shah Manisha. 2013. "Can Mistargeting Destroy Social Capital and Stimulate Crime? Evidence from a Cash Transfer Program in Indonesia." *Economic Development and Cultural Change* 62 (2):381-415.

Campaniello, Nadia, Gray Rowena, and Mastrobuoni Giovanni. 2016. "Returns to Education in Criminal Organizations: Did Going to College Help Michael Corleone?" *Economics of Education Review* 54:242-58.

Carvalho, José Raimundo, and Lavor Sylvia Cristina. 2008. "Repeat Property Criminal Victimization and Income Inequality in Brazil." *Economica (Selecta)* 9 (4):87–110.

Cerdan-Infantes, Pedro, Makarova Yulia, Al-Samarrai Samer, and Chen Dandan. 2013. "Spending More or Spending Better: Improving Education Financing in Indonesia." *Policy brief. Washington DC: World Bank.*

Chakrabarty, Sayan, Grote Ulrike, and Lüchters Guido. 2011. "Does Social Labelling Encourage Child Schooling and Discourage Child Labour in Nepal?" *International Journal of Educational Development* 31 (5):483-9.

Chamarbagwala, Rubiana. 2008. "Regional Returns to Education, Child Labour and Schooling in India." *The Journal of Development Studies* 44 (2):233-57.

Chan, K, Shum Connie, and Wright David. 1997. "Class Attendance and Student Performance in Principles of Finance." *Financial Practice and Education* 7 (2):58-65.

Chaudhuri, Sarbajit. 2004. "Incidence of Child Labour, Free Education Policy, and Economic Liberalisation in a Developing Economy." *The Pakistan development review* 43 (1):1-25.

Chevalier, Arnaud, Harmon Colm, O'Sullivan Vincent, and Walker Ian. 2013. "The Impact of Parental Income and Education on the Schooling of Their Children." *IZA Journal of Labor Economics* 2 (1):8.

Chioda, Laura, De Mello João MP, and Soares Rodrigo R. 2016. "Spillovers from Conditional Cash Transfer Programs: Bolsa Família and Crime in Urban Brazil." *Economics of Education Review* 54:306-20.

Christofides, Louis N, Hoy Michael, Milla Joniada, and Stengos Thanasis. 2015. "Grades, Aspirations and Post-Secondary Education Outcomes." *Canadian Journal of Higher Education* 45 (1):48-82.

Cigno, Alessandro, and Rosati Furio Camillo. 2005. *The Economics of Child Labour*. Cambridge: Oxford University Press.

Cinnirella, Francesco, and Streb Jochen. 2017. "The Role of Human Capital and Innovation in Economic Development: Evidence from Post-Malthusian Prussia." *Journal of economic growth* 22 (2):193-227.

Committee, Central Population and Housing Census Steering. 2010. *The 2009 Vietnam Population and Housing Census: Major Findings*. 6. 2010 ed. Hanoi.

Cook, Michael D., and Evans William N. 2000. "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *Journal of Labor Economics* 18 (4):729-54.

Cook, Philip J, and Kang Songman. 2016. "Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation." *American Economic Journal: Applied Economics* 8 (1):33-57.

Creed, Peter A, Conlon Elizabeth G, and Zimmer-Gembeck Melanie J. 2007. "Career Barriers and Reading Ability as Correlates of Career Aspirations and Expectations of Parents and Their Children." *Journal of Vocational Behavior* 70 (2):242-58.

Dammert, Ana C. 2010. "Siblings, Child Labor, and Schooling in Nicaragua and Guatemala." *Journal of Population Economics* 23 (1):199-224.

Dang, Hai-Anh. 2012. *Vietnam: A Widening Poverty Gap for Ethnic Minorities*. Edited by Gillette Hall and Harry Patrinos, *Indigenous Peoples, Poverty and Development*: Cambridge University Press.

Dang, Hai-Anh H, and Glewwe Paul W. 2017. "Well Begun, but Aiming Higher: A Review of Vietnam's Education Trends in the Past 20 Years and Emerging Challenges." *World Bank Policy Research Paper # 8112*.

Dang, Hai-Anh, and Rogers F Halsey. 2013. "The Decision to Invest in Child Quality over Quantity: Household Size and Household Investment in

Education in Vietnam." *World Bank Policy Research Working Paper* (6487).

Davis-Kean, Pamela E. 2005. "The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment." *Journal of family psychology* 19 (2):294.

De Juan, Alexander, Pierskalla Jan H, and Vüllers Johannes. 2015. "The Pacifying Effects of Local Religious Institutions: An Analysis of Communal Violence in Indonesia." *Political Research Quarterly* 68 (2):211-24.

de Walque, Damien. 2010. "Education, Information, and Smoking Decisions Evidence from Smoking Histories in the United States, 1940–2000." *Journal of Human Resources* 45 (3):682-717.

Dee, Thomas S. 2004. "Are There Civic Returns to Education?" *Journal of Public Economics* 88 (9-10):1697-720.

Deming, David J. 2011. "Better Schools, Less Crime?" *The Quarterly Journal of Economics* 126 (4):2063-115.

Doan, Tinh Thanh. 2011. "Labour Market Returns to Higher Education in Vietnam." *Economics Discussion Paper* (2011-4).

Dorman, Peter. 2008. *Child Labour, Education and Health: A Review of the Literature*. Geneva: ILO.

Du, Yang. 2013. "Do the Recent Labor Market Changes Negatively Affect Schooling." *China & World Economy* 21 (2):38-55.

Edmonds, Eric V. 2005. "Does Child Labor Decline with Improving Economic Status?" *Journal of human resources* 40 (1):77-99.

Edmonds, Eric V. 2006a. "Child Labor and Schooling Responses to Anticipated Income in South Africa." *Journal of Development Economics* 81 (2):386-414.

Edmonds, Eric V. 2006b. "Understanding Sibling Differences in Child Labor." *Journal of Population Economics* 19 (4):795-821.

Edmonds, Eric V. 2008. "Child Labor." In *Handbook of Development Economics*, edited by Schultz Paul and John Strauss, 3607-709. Amsterdam: Elsevier.

Edmonds, Eric V., and Turk Carrie. 2002. "Child Labor in Transition in Vietnam." *World Bank Policy Research Working Paper* (2774).

Ejrnæs, Mette, and Pörtner Claus C. 2004. "Birth Order and the Intrahousehold Allocation of Time and Education." *Review of Economics and Statistics* 86 (4):1008-19.

Emerson, Patrick M, and Souza Andre Portela. 2002. "Bargaining over Sons and Daughters: Child Labor, School Attendance and Intra-Household Gender Bias in Brazil." In.: Vanderbilt University, Economics Department.

Emerson, Patrick M, and Souza Andre Portela. 2008. "Birth Order, Child Labor, and School Attendance in Brazil." *World Development* 36 (9):1647-64.

Ersado, Lire. 2002. "Child Labor and School Decisions in Urban and Rural Areas: Cross Country Evidence." *FCND Discussion Paper* No. 145.

Ersado, Lire. 2005. "Child Labor and Schooling Decisions in Urban and Rural Areas: Comparative Evidence from Nepal, Peru, and Zimbabwe." *World Development* 33 (3):455-80.

Ewert, Stephanie, Sykes Bryan L., and Pettit Becky. 2014. "The Degree of Disadvantage: Incarceration and Inequality in Education." *The ANNALS of the American Academy of Political and Social Science* 651 (1):24-43.

Fan, Xitao, and Chen Michael. 2001. "Parental Involvement and Students' Academic Achievement: A Meta-Analysis." Educational psychology review 13 (1):1-22.

Favara, Marta. 2017. "Do Dreams Come True? Aspirations and Educational Attainments of Ethiopian Boys and Girls." *Journal of African Economies* 26 (5):561-83.Filmer, Deon, and Scott Kinnon. 2012. "Assessing Asset Indices." *Demography* 49 (1):359-92.

Fitrani, Fitria, Hofman Bert, and Kaiser Kai. 2005. "Unity in Diversity? The Creation of New Local Governments in a Decentralising Indonesia." *Bulletin of Indonesian Economic Studies* 41 (1):57-79.

Ford, Margot. 2013. "Achievement Gaps in Australia: What Naplan Reveals About Education Inequality in Australia." *Race ethnicity and education* 16 (1):80-102.

Froiland, John Mark, Peterson Aubrey, and Davison Mark L. 2013. "The Long-Term Effects of Early Parent Involvement and Parent Expectation in the USA." *School Psychology International* 34 (1):33-50.

Gennaioli, Nicola, La Porta Rafael, Lopez-de-Silanes Florencio, and Shleifer Andrei. 2012. "Human Capital and Regional Development." *The Quarterly journal of economics* 128 (1):105-64.

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

Greenstone, Michael, and Gallagher Justin. 2008. "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program." *The Quarterly Journal of Economics* 123 (3):951-1003.

Grimard, Franque, and Parent Daniel. 2007. "Education and Smoking: Were Vietnam War Draft Avoiders Also More Likely to Avoid Smoking?" *Journal of Health Economics* 26 (5):896-926.

GSO. 2010a. "The 2009 Vietnam Population and Housing Census: Completed Results." In. Hanoi, Vietnam: Central Populatin and Housing Census Steering Committe.

GSO. 2010b. *The 2009 Vietnam Population and Housing Census: Major Findings*. Hanoi: Central Population and Housing Census Steering Committee, General Statistics Office.

GSO. 2011. "Result of the Viet Nam Household Living Standards Survey 2010." In. Hanoi, Vietnam: Statistical Publishing House, General Statistics Office.

GSO. 2016. *Statistical Yearbook of Vietnam*. Hanoi, Vietnam: Statistic Publishing House.

Gunnarsson, Victoria, Orazem Peter F., and Sánchez Mario A. 2006. "Child Labor and School Achievement in Latin America." *The World Bank Economic Review* 20 (1):31-54.

Hacker, Douglas J, Bol Linda, Horgan Dianne D, and Rakow Ernest A. 2000. "Test Prediction and Performance in a Classroom Context." *Journal of Educational Psychology* 92 (1):160.

Hahn, Jinyong, Todd Petra, and Van der Klaauw Wilbert. 2001. "Identification and Estimation of Treatment Effects with a Regression -Discontinuity *Econometrica* 69 (1):201-9.

Haile, Demewoz, Nigatu Dabere, Gashaw Ketema, and Demelash Habtamu. 2016. "Height for Age Z Score and Cognitive Function Are Associated with Academic Performance among School Children Aged 8-11 Years Old." *Archives of public health* 74 (1):1.

Hanushek, Eric A, and Woessmann Ludger. 2008. "The Role of Cognitive Skills in Economic Development." *Journal of economic literature* 46 (3):607-68.

Heady, Christopher. 2003. "The Effect of Child Labor on Learning Achievement." *World Development* 31 (2):385-98.

Hendri, Davy, and Muharja Fajri. 2013. "Income Inequality Impact on Crime in Indonesia: Static and Dynamic Analysis During 2007-2011." *Working paper*.

Hjalmarsson, Randi. 2008. "Criminal Justice Involvement and High School Completion." *Journal of Urban Economics* 63 (2):613-30.

Hjalmarsson, Randi, Holmlund Helena, and Lindquist Matthew J. 2015. "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro -Data." *The Economic Journal* 125 (587):1290-326.

Hong, Seo Yeon. 2013. "Three Essays on Child Labor and Education in Developing Countries." The Pardee RAND Graduate School

Hou, Xiaohui. 2010. "Wealth: Crucial but Not Sufficient–Evidence from Pakistan on Economic Growth, Child Labour and Schooling." *The Journal of Development Studies* 46 (3):439-65.

Hussin, Nasimah, and Zawawi Majdah. 2012. "Preventing Criminal Victimization through Community Education: An Islamic Formula." *Procedia-Social and Behavioral Sciences* 68:855-64.

Ilahi, Nadeem. 2001. "Children's Work and Schooling: Does Gender Matter?" *Worl Bank Policy Research Working Papers* (2745).

Imai, Katsushi S, Gaiha Raghav, and Kang Woojin. 2011. "Poverty, Inequality and Ethnic Minorities in Vietnam." *International review of applied economics* 25 (3):249-82.

Islam, Asadul, and Choe Chongwoo. 2013. "Child Labor and Schooling Responses to Access to Microcredit in Rural Bangladesh." *Economic Inquiry* 51 (1):46-61.

Jacobson, Jerry Owen, Robinson Paul, and Bluthenthal Ricky N. 2007. "A Multilevel Decomposition Approach to Estimate the Role of Program Location and Neighborhood Disadvantage in Racial Disparities in Alcohol Treatment Completion." *Social Science & Medicine* 64 (2):462-76.

Jonathan, London. 2011. "Contemporary Vietnam's Education System: Historical Roots, Current Trends." *Education in Vietnam*:1-57.

Jonck, Petronella, Goujon Anne, Testa Maria Rita, and Kandala John. 2015. "Education and Crime Engagement in South Africa: A National and Provincial Perspective." *International Journal of Educational Development* 45:141-51.

Kambhampati, Uma. 2009. "Child Schooling and Work Decisions in India: The Role of Household and Regional Gender Equity." *Feminist Economics* 15 (4):77-112.

Kelly, Morgan. 2000. "Inequality and Crime." *The Review of Economics and Statistics* 82 (4):530-9.

Kis-Katos, Krisztina, and Sparrow Robert. 2011. "Child Labor and Trade Liberalization in Indonesia." *Journal of Human Resources* 46 (4):722-49.

Kruger, Diana I. 2007. "Coffee Production Effects on Child Labor and Schooling in Rural Brazil." *Journal of Development Economics* 82 (2):448-63.

Kumar, Alok. 2015. "One More Year of Schooling or Work: Grade-Transition and Child Labor in Rural Bangladesh." *Journal of developing areas* 49 (3):27-48.

Leigh, Andrew, and Gong Xiaodong. 2009. "Estimating Cognitive Gaps between Indigenous and Non -Indigenous Australians." *Economic Record* 85 (2):239-61.

Lewis, Blane. 2014. "Twelve Years of Fiscal Decentralization: A Balance Sheet." *Regional dynamics in a decentralized Indonesia*:135-55.

Lewis, Blane D. 2017. "Does Local Government Proliferation Improve Public Service Delivery? Evidence from Indonesia." *Journal of Urban Affairs* 39 (8):1047-65.

Lillydahl, Jane H. 1990. "Academic Achievement and Part-Time Employment of High School Students." *The journal of economic education* 21 (3):307-16.

Lima, Luiz Renato, Mesquita Shirley, and Wanamaker Marianne. 2015. "Child Labor and the Wealth Paradox: The Role of Altruistic Parents." *Economics letters* 130:80-2.

Liu, Amy YC. 2000. *Children at Work: Do Regions Matter and Why? A Case Study in Vietnam, Working Papers*. Australian National University: Asia Pacific Press.

Liu, Amy YC. 2004. "Flying Ducks? Girls' Schooling in Rural Vietnam: A Revisit." *Asian Economic Journal* 18 (3):293-318.

Lochner, Lance. 2011. "Non-Production Benefits of Education: Crime, Health, and Good Citizenship." In.: National Bureau of Economic Research.

Lochner, Lance, and Moretti Enrico. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *The American Economic Review* 94 (1):155-89.

Machin, Stephen, Marie Olivier, and Vujić Sunčica. 2011. "The Crime Reducing Effect of Education." *The Economic Journal* 121 (552):463-84.

Machin, Stephen, Marie Olivier, and Vujić Sunčica. 2012. "Youth Crime and Education Expansion." *German Economic Review* 13 (4):366-84.

Maitra, Pushkar, and Ray Ranjan. 2002. "The Joint Estimation of Child Participation in Schooling and Employment: Comparative Evidence from Three Continents." *Oxford Development Studies* 30 (1):41-62.

Malik, Bijaya Kumar. 2013. "Child Schooling and Child Work in India: Does Poverty Matter?" *International Journal of Child Care and Education Policy* 7 (1):80-101.

Mansoob Murshed, Syed, Zulfan Tadjoeddin Mohammad, and Chowdhury Anis. 2009. "Is Fiscal Decentralization Conflict Abating? Routine Violence and District Level Government in Java, Indonesia." *Oxford Development Studies* 37 (4):397-421.

McAdams, John M. 2016. "The Effect of School Starting Age Policy on Crime: Evidence from Us Microdata." *Economics of Education Review* 54:227-41.

McCaig, Brian, Benjamin Dwayne, and Brandt Loren. 2015. "Growth with Equity: Income Inequality in Vietnam, 2002-12." In.

Melick, Matthew D. 2003. "The Relationship between Crime and Unemployment." The Park Place Economist 11 (1):30-6.

Milligan, Kevin, Moretti Enrico, and Oreopoulos Philip. 2004. "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom." *Journal of public Economics* 88 (9-10):1667-95.

MOEC. 2013. *Management of National Education in 2011/2012 at a Glance*. Jakarta: Center for Educational and Cultural Data and Statistics, MoEC.

MOET. 2015. "Vietnam National Education for All 2015 Review." In. Ha Noi: Ministry of Education and Training.

MOET, GSO, and UIS. 2016. "Education Financing in Viet Nam 2009-2013." In. Hanoi: Ministry of Education and Training; General Statistical Office; UNESCO Institute for Statistics.

Montgomery, Mark R, Gragnolati Michele, Burke Kathleen A, and Paredes Edmundo. 2000. "Measuring Living Standards with Proxy Variables." *Demography* 37 (2):155-74.

MPC. 2017. "Integrated Public Use Microdata Series, International: Version 6.5 [Dataset]." In. Minneapolis: University of Minnesota.

Mukherjee, Diganta, and Das Saswati. 2008. "Role of Parental Education in Schooling and Child Labour Decision: Urban India in the Last Decade." *Social Indicators Research* 89 (2):305-22.

Nguyen, D, Parker L, Doan M, and Brennan L. 2012. Decision Making Dynamics within the Vietnamese Family Unit. Paper presented at the in Proceedings of ANZMAC 2012.

Nguyen, Ngoc P. 2008. "An Assessment of the Young Lives Sampling Approach in Vietnam." *Young Lives Technical Note No. 4*.

Nørlund, Irene, Gates Carolyn, and Vu Cao Dam. 1995. *Vietnam in a Changing World*. Vol. 17: Psychology Press.

OECD, and ADB. 2015. *Education in Indonesia: Rising to the Challenge*. Paris: OECD Publishing.

Oreopoulos, Philip. 2006. "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter." *American Economic Review* 96 (1):152-75.

Papps, Kerry, and Winkelmann Rainer. 2000. "Unemployment and Crime: New Evidence for an Old Question." New Zealand Economic Papers 34 (1):53-71.

Parinduri, Rasyad A. 2014. "Do Children Spend Too Much Time in Schools? Evidence from a Longer School Year in Indonesia." *Economics of Education Review* 41:89-104.

Paxson, Christina, and Schady Norbert. 2007. "Cognitive Development among Young Children in Ecuador the Roles of Wealth, Health, and Parenting." *Journal of Human resources* 42 (1):49-84.

Pham, Hung T, and Reilly Barry. 2009. "Ethnic Wage Inequality in Vietnam." *International Journal of Manpower* 30 (3):192-219.

Pierskalla, Jan H, and Sacks Audrey. 2017. "Unpacking the Effect of Decentralized Governance on Routine Violence: Lessons from Indonesia." *World Development* 90:213-28.

Powers, Daniel A, Yoshioka Hirotoshi, and Yun Myeong-Su. 2011. "Mvdcmp: Multivariate Decomposition for Nonlinear Response Models." *Stata Journal* 11 (4):556-76.

Psacharopoulos, George. 1997. "Child Labor Versus Educational Attainment Some Evidence from Latin America." *Journal of population economics* 10 (4):377-86.

Quyen, Bui Thai. 2011. "School Dropout Trends in Vietnam from 1998 to 2006." *Education in Vietnam*:152-70.

Raphael, Steven, and Winter-Ebmer Rudolf. 2001. "Identifying the Effect of Unemployment on Crime." *The Journal of Law and Economics* 44 (1):259-83.

Rasbash, Jon. 2008. *Module 4: Multilevel Structures and Classifications*. University of Bristol: Centre for Multilevel Modelling.

Ray, Ranjan. 2000. "Child Labor, Child Schooling, and Their Interaction with Adult Labor: Empirical Evidence for Peru and Pakistan." *The World Bank Economic Review* 14 (2):347-67.

Ray, Ranjan. 2002. "The Determinants of Child Labour and Child Schooling in Ghana." *Journal of African economies* 11 (4):561-90.

Ray, Ranjan, and Lancaster Geoffrey. 2005. "The Impact of Children's Work on Schooling: Multi-Country Evidence." *International Labour Review* 144 (2):189-210.

Rew, W Joshua. 2008. "Provincial, Ethnic, and Gender Disparities in Education: A Descriptive Study of Vietnam." In *Inequality in Education*, 307-23. Springer.

Ridao-Cano, Cristobal. 2001. "Child Labor and Schooling in a Low Income Rural Economy." *Journal of Economic Literature*.

Rolleston, Caine , James Zoe , Laure Pasquier Doumer, and Tam Tran Ngo Thi Minh 2013. "Making Progress: Report of the Young Lives School Survey in Vietnam." *Young Lives Working Paper No. 100*.

Roodman, David. 2009. "How to Do Xtabond2: An Introduction to Difference and System Gmm in Stata." *Stata Journal* 9 (1):86-136.

Roodman, David. 2011. "Fitting Fully Observed Recursive Mixed-Process Models with Cmp." *Stata Journal* 11 (2):159-206.

Rosati, Furio Camillo, and Rossi Mariacristina. 2001. "Children's Working Hours, School Enrolment and Human Capital Accumulation: Evidence from Pakistan and Nicaragua." *Understanding Children's Work Project Working Paper Series* (43861).

Sakellariou, Chris. 2008. "Peer Effects and the Indigenous/Non-Indigenous Early Test-Score Gap in Peru." *Education Economics* 16 (4):371-90.

Schoon, Ingrid, Parsons Samantha, and Sacker Amanda. 2004. "Socioeconomic Adversity, Educational Resilience, and Subsequent Levels of Adult Adaptation." *Journal of Adolescent Research* 19 (4):383-404.

Seid, Yared. 2013. "Essays in Children's Time Allocation and Age at Primary School Enrollment." PhD Dissertation in Economics, Georgia State University.

Shafiq, M Najeeb. 2007. "Household Schooling and Child Labor Decisions in Rural Bangladesh." *Journal of Asian Economics* 18 (6):946-66.

Silles, Mary A. 2009. "The Causal Effect of Education on Health: Evidence from the United Kingdom." *Economics of Education review* 28 (1):122-8.

Skoufias, Emmanuel. 1993. "Labor Market Opportunities and Intrafamily Time Allocation in Rural Households in South Asia." *Journal of Development Economics* 40 (2):277-310.

Skoufias, Emmanuel. 1994. "Market Wages, Family Composition and the Time Allocation of Children in Agricultural Households." *The Journal of Development Studies* 30 (2):335-60.

Soares, Rodrigo R, Kruger Diana, and Berthelon Matias. 2012. "Household Choices of Child Labor and Schooling a Simple Model with Application to Brazil." *Journal of Human Resources* 47 (1):1-31.

Stanca, Luca. 2006. "The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics." *The Journal of Economic Education* 37 (3):251-66.

Strand, Steve. 2011. "The Limits of Social Class in Explaining Ethnic Gaps in Educational Attainment." *British Educational Research Journal* 37 (2):197-229.

Suryadarma, Daniel, Suryahadi Asep, Sumarto Sudarno, and Rogers F Halsey. 2006. "Improving Student Performance in Public Primary Schools in Developing Countries: Evidence from Indonesia." *Education Economics* 14 (4):401-29.

Tadjoeddin, Mohammad Zulfan, and Murshed Syed Mansoob. 2007. "Socio-Economic Determinants of Everyday Violence in Indonesia: An Empirical Investigation of Javanese Districts, 1994—2003." *Journal of Peace Research* 44 (6):689-709.

TI. Transparency International. https://www.transparency.org/gcb2013/country?country=indonesia.

Van de Walle, Dominique, and Gunewardena Dileni. 2001. "Sources of Ethnic Inequality in Viet Nam." *Journal of Development Economics* 65 (1):177-207.

Van Der Linden, Wim J, and Hambleton Ronald K. 1997. "Item Response Theory: Brief History, Common Models, and Extensions." *Handbook of modern item response theory*:1-28.

Ward, Shannon, and Williams Jenny. 2015. "Does Juvenile Delinquency Reduce Educational Attainment?" *Journal of Empirical Legal Studies* 12 (4):716-56.

WB. 2009. *Country Social Analysis: Ethnicity and Development in Vietnam*. Washington, DC: World Bank.

Webbink, Dinand, Koning Pierre, Vujić Sunčica, and Martin Nicholas G. 2012. "Why Are Criminals Less Educated Than Non-Criminals? Evidence from a Cohort of Young Australian Twins." *The Journal of Law, Economics, & Organization* 29 (1):115-44.

WHO. 1997. *Global Database on Child Growth and Malnutrition*, . Geneva: WHO.

Windmeijer, Frank. 2005. "A Finite Sample Correction for the Variance of Linear Efficient Two-Step Gmm Estimators." *Journal of econometrics* 126 (1):25-51.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*: MIT press.

YL. 2002. "Young Lives Conceptual Framework." *UK Data Archive Study Number 5307 - Young Lives: an International Study of Childhood Poverty: Round 1*.

Zabaleta, Mariela Buonomo. 2011. "The Impact of Child Labor on Schooling Outcomes in Nicaragua." *Economics of Education Review* 30 (6):1527.

Zhang, Yanyan, Haddad Eileen, Torres Bernadeth, and Chen Chuansheng. 2011. "The Reciprocal Relationships among Parents' Expectations, Adolescents' Expectations, and Adolescents' Achievement: A Two-Wave Longitudinal Analysis of the Nels Data." *Journal of youth and adolescence* 40 (4):479-89.

Appendix

Table A1. Definition of variables in Young Lives surveys

Variables	Definition
Hours of study	Daily hours of study
Hours of economic work	Daily hours of economic work
Hours of housework	Daily hours of housework
Parental aspirations	Years of school that parents wish their child to complete
Younger Cohort	Dummy = 1 if a child belongs to younger cohort
Male	Dummy = 1 if a child is male
Child's age in month	Child's age in month
BMI-for-age z-score (-1)	BMI-for-age z-score in previous survey
Birth order	Child’s birth order
Child's highest grade	Child's highest grade completed
Adjusted PPVT	IRT-Adjusted PPVT
Non-Kinh child	Dummy = 1 if child is non-Kinh ethnicity
Number of children aged 0-5	Number of children aged 0-5
Education of caregiver	Years of school of the caregiver
Main caregiver is a parent	Dummy = 1 if the caregiver is a parent
Wealth index	Wealth index
North East	Dummy = 1 if the region is North East
Red River Delta	Dummy = 1 if the region is Red River Delta
South Central Coast	Dummy = 1 if the region is South Central Coast
Mekong River Delta	Dummy = 1 if the region is South Central Coast
Ln(average local wage)	Logarithm of the average local wage

Table A2. Mixed process models: hours of study, economic work and chore

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
Parental aspirations	**0.199*** (0.021)**	**-0.117** (0.046)**	**-0.023 (0.018)**	
Younger Cohort	0.834*** (0.078)	-0.511*** (0.183)	-0.052 (0.067)	-2.532*** (0.082)
Male	-0.231*** (0.066)	0.157 (0.156)	-0.519*** (0.056)	-1.210*** (0.066)
Child's age in month	-0.054*** (0.010)	0.025 (0.023)	0.016* (0.008)	-0.032*** (0.009)
BMI-for-age z-score (-1)	0.001 (0.029)	-0.043 (0.074)	0.022 (0.025)	0.095*** (0.028)
Birth order	-0.022 (0.033)	0.106 (0.073)	-0.077*** (0.028)	0.040 (0.032)
Child's highest grade	0.539*** (0.047)	-0.181* (0.098)	0.047 (0.040)	0.416*** (0.042)
Adjusted PPVT	0.001 (0.002)	-0.005 (0.004)	0.001 (0.002)	0.001 (0.002)

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
Non-Kinh child	-0.094	0.974***	-0.030	0.151
	(0.139)	(0.289)	(0.117)	(0.132)
Number of children aged 0-5	-0.077	-0.431***	0.723***	-0.099
	(0.066)	(0.157)	(0.056)	(0.063)
Education of caregiver	0.032***	-0.051*	-0.030***	0.063***
	(0.011)	(0.029)	(0.010)	(0.011)
Main caregiver is a parent	0.147	0.850**	-0.058	0.132
	(0.156)	(0.424)	(0.133)	(0.148)
Wealth index	0.549*	-0.624	-0.628**	2.978***
	(0.317)	(0.749)	(0.271)	(0.292)
North East	-0.701***	2.928***	0.946***	3.462***
	(0.136)	(0.347)	(0.117)	(0.141)
Red River Delta	0.363***	1.542***	1.173***	3.631***
	(0.109)	(0.311)	(0.096)	(0.128)
South Central Coast	-0.598***	1.815***	0.974***	3.035***
	(0.117)	(0.316)	(0.102)	(0.124)
Mekong River Delta	-0.579***	0.931***	0.726***	4.495***
	(0.119)	(0.334)	(0.104)	(0.140)

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
Ln(average local wage)				5.797***
				(0.117)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A3. Mixed process models: hours of study, economic work and chore for boys

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
Parental aspirations	**0.173*** (0.032)**	**-0.166** (0.070)**	**-0.003 (0.029)**	
Younger Cohort	0.846*** (0.114)	-0.577** (0.268)	0.025 (0.102)	-2.113*** (0.119)
Child's age in month	-0.053*** (0.014)	0.018 (0.033)	0.013 (0.012)	-0.043*** (0.014)
BMI-for-age z-score (-1)	-0.020 (0.038)	-0.021 (0.100)	0.033 (0.035)	0.044 (0.038)
Birth order	-0.016 (0.047)	0.166 (0.104)	-0.116*** (0.043)	0.087* (0.047)
Child's highest grade	0.579*** (0.064)	-0.260** (0.132)	-0.065 (0.057)	0.606*** (0.057)
Adjusted PPVT	0.002 (0.003)	0.001 (0.007)	-0.000 (0.002)	0.001 (0.003)
Non-Kinh child	-0.030 (0.194)	1.021** (0.405)	-0.050 (0.171)	0.413** (0.192)
Number of children aged 0-5	-0.082	-0.136	0.721***	-0.064

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
	(0.097)	(0.228)	(0.086)	(0.096)
Education of caregiver	0.036**	-0.062	-0.016	0.059***
	(0.017)	(0.042)	(0.015)	(0.016)
Main caregiver is a parent	0.306	1.156*	-0.098	0.173
	(0.229)	(0.648)	(0.205)	(0.226)
Wealth index	1.088**	-1.167	-0.695*	3.169***
	(0.444)	(1.074)	(0.397)	(0.427)
North East	-0.521***	3.085***	1.107***	3.146***
	(0.192)	(0.493)	(0.176)	(0.210)
Red River Delta	0.409***	1.313***	1.431***	3.513***
	(0.155)	(0.451)	(0.145)	(0.191)
South Central Coast	-0.509***	2.192***	1.175***	3.024***
	(0.163)	(0.447)	(0.152)	(0.181)
Mekong River Delta	-0.377**	0.538	0.897***	5.368***
	(0.169)	(0.492)	(0.158)	(0.231)
Ln(average local wage)				5.754***
				(0.180)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A4. Mixed process models: hours of study, economic work and chore for girls

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
Parental aspirations	**0.240*** (0.027)**	**-0.141*** (0.054)**	**-0.044** (0.021)**	
Younger Cohort	0.835*** (0.108)	-0.492** (0.242)	-0.125 (0.088)	-3.357*** (0.103)
Child's age in month	-0.052*** (0.014)	0.026 (0.031)	0.017 (0.011)	-0.013 (0.011)
BMI-for-age z-score (-1)	0.019 (0.045)	-0.058 (0.109)	0.010 (0.037)	0.135*** (0.036)
Birth order	-0.026 (0.047)	0.020 (0.100)	-0.033 (0.038)	-0.013 (0.038)
Child's highest grade	0.470*** (0.070)	-0.010 (0.146)	0.197*** (0.057)	0.065 (0.056)
Adjusted PPVT	0.001 (0.003)	-0.009 (0.006)	0.001 (0.002)	-0.000 (0.002)
Non-Kinh child	-0.166 (0.200)	0.802** (0.403)	0.013 (0.161)	0.034 (0.161)
Number of children aged 0-5	-0.066	-0.766***	0.726***	-0.162**

	Hours of study	Hours of economic work	Hours of chore	Parental aspirations
	(0.090)	(0.213)	(0.072)	(0.072)
Education of caregiver	0.027*	-0.037	-0.040***	0.062***
	(0.016)	(0.038)	(0.013)	(0.012)
Main caregiver is a parent	-0.006	0.791	-0.022	0.034
	(0.214)	(0.550)	(0.173)	(0.172)
Wealth index	-0.146	0.093	-0.664*	2.654***
	(0.456)	(1.017)	(0.370)	(0.356)
North East	-0.938***	2.828***	0.805***	4.296***
	(0.192)	(0.476)	(0.156)	(0.169)
Red River Delta	0.301*	1.754***	0.931***	4.484***
	(0.154)	(0.421)	(0.127)	(0.157)
South Central Coast	-0.686***	1.401***	0.767***	3.654***
	(0.169)	(0.438)	(0.139)	(0.155)
Mekong River Delta	-0.787***	1.225***	0.565***	4.351***
	(0.168)	(0.445)	(0.139)	(0.159)
Ln(average local wage)				7.024***
				(0.150)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A5. GMM estimation: hours of study, economic work and chore, all children

	Hours of study	Hours of economic work	Hours of chore
Parental aspirations	**0.199*** **	**-0.080*** **	**-0.030* **
	(0.027)	**(0.020)**	**(0.016)**
Younger Cohort	0.834***	-0.210***	-0.037
	(0.081)	(0.049)	(0.051)
Male	-0.231***	0.072*	-0.380***
	(0.066)	(0.040)	(0.044)
Child's age in month	-0.054***	0.016**	0.011
	(0.010)	(0.007)	(0.007)
BMI-for-age z-score (-1)	0.001	0.018	0.017
	(0.029)	(0.016)	(0.019)
Birth order	-0.022	0.034	-0.058**
	(0.036)	(0.025)	(0.023)
Child's highest grade	0.539***	-0.148**	0.041
	(0.071)	(0.062)	(0.043)
Adjusted PPVT	0.001	-0.001	0.000
	(0.002)	(0.001)	(0.001)
Non-Kinh child	-0.094	0.470***	0.009
	(0.147)	(0.116)	(0.107)
Number of children aged 0-5	-0.077	-0.096**	0.634***
	(0.067)	(0.049)	(0.054)
Education of caregiver	0.032***	-0.003	-0.022***
	(0.011)	(0.006)	(0.007)
Main caregiver is a parent	0.147	0.108	-0.079
	(0.163)	(0.080)	(0.116)
Wealth index	0.549	0.078	-0.368*
	(0.345)	(0.210)	(0.211)
North East	-0.701***	0.619***	0.580***
	(0.128)	(0.085)	(0.086)
Red River Delta	0.363***	0.226***	0.721***
	(0.108)	(0.046)	(0.065)
South Central Coast	-0.598***	0.276***	0.646***
	(0.118)	(0.059)	(0.081)
Mekong River Delta	-0.579***	0.087	0.461***

	Hours of study	Hours of economic work	Hours of chore
	(0.119)	(0.054)	(0.079)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A6. GMM estimation: hours of study, economic work and chore for boys

	Hours of study	Hours of economic work	Hours of chore
Parental aspirations	**0.173*** **	**-0.107*** **	**-0.013**
	(0.037)	**(0.031)**	**(0.021)**
Younger Cohort	0.846***	-0.204***	-0.004
	(0.113)	(0.072)	(0.073)
Child's age in month	-0.053***	0.015	0.007
	(0.014)	(0.009)	(0.009)
BMI-for-age z-score (-1)	-0.020	0.016	0.024
	(0.039)	(0.022)	(0.025)
Birth order	-0.016	0.073*	-0.081**
	(0.051)	(0.038)	(0.033)
Child's highest grade	0.579***	-0.170*	-0.052
	(0.099)	(0.088)	(0.057)
Adjusted PPVT	0.002	0.000	-0.000
	(0.003)	(0.002)	(0.002)
Non-Kinh child	-0.030	0.526***	-0.041
	(0.196)	(0.174)	(0.140)
Number of children aged 0-5	-0.082	-0.043	0.612***
	(0.097)	(0.083)	(0.080)
Education of caregiver	0.036**	-0.010	-0.012
	(0.016)	(0.010)	(0.010)
Main caregiver is a parent	0.306	0.134	-0.135
	(0.257)	(0.129)	(0.184)
Wealth index	1.088**	0.047	-0.348
	(0.480)	(0.321)	(0.282)
North East	-0.521***	0.698***	0.628***
	(0.172)	(0.133)	(0.117)
Red River Delta	0.409***	0.186***	0.796***
	(0.152)	(0.066)	(0.089)
South Central Coast	-0.509***	0.347***	0.720***
	(0.161)	(0.089)	(0.112)
Mekong River Delta	-0.377**	0.005	0.505***
	(0.165)	(0.082)	(0.106)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A7. GMM estimation: hours of study, economic work and chore for girls

	Hours of study	Hours of economic work	Hours of chore
Parental aspirations	**0.240*** **	**-0.074*** **	**-0.051** **
	(0.038)	**(0.022)**	**(0.021)**
Younger Cohort	0.835***	-0.229***	-0.075
	(0.114)	(0.063)	(0.071)
Child's age in month	-0.052***	0.016*	0.012
	(0.015)	(0.009)	(0.010)
BMI-for-age z-score (-1)	0.019	0.028	0.006
	(0.044)	(0.021)	(0.032)
Birth order	-0.026	-0.014	-0.026
	(0.052)	(0.030)	(0.031)
Child's highest grade	0.470***	-0.102	0.174***
	(0.096)	(0.075)	(0.059)
Adjusted PPVT	0.001	-0.002	-0.000
	(0.003)	(0.002)	(0.002)
Non-Kinh child	-0.166	0.366**	0.076
	(0.218)	(0.148)	(0.160)
Number of children aged 0-5	-0.066	-0.158***	0.651***
	(0.092)	(0.049)	(0.073)
Education of caregiver	0.027*	0.003	-0.030***
	(0.015)	(0.008)	(0.011)
Main caregiver is a parent	-0.006	0.138*	-0.026
	(0.204)	(0.082)	(0.145)
Wealth index	-0.146	0.173	-0.445
	(0.495)	(0.254)	(0.318)
North East	-0.938***	0.559***	0.533***
	(0.186)	(0.105)	(0.125)
Red River Delta	0.301*	0.276***	0.631***
	(0.154)	(0.064)	(0.092)
South Central Coast	-0.686***	0.207***	0.541***
	(0.174)	(0.076)	(0.117)
Mekong River Delta	-0.787***	0.140**	0.402***
	(0.172)	(0.069)	(0.116)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A8. Mixed process models: Maths score, hours of study, economic work, and parental aspirations

	Maths score	Hours of study	Hours of economic work	Parental aspirations
Hours of study	**0.859*** **			
	(0.120)			
Hours of economic work	**-0.076**			
	(0.058)			
Parental aspirations		**0.204*** **	**-0.127*** **	
		(0.021)	**(0.045)**	
Male	0.118	-0.218***		-1.217***
	(0.079)	(0.064)		(0.066)
Younger Cohort	-3.641***	0.846***	-0.550***	-2.533***
	(0.122)	(0.078)	(0.182)	(0.082)
Child's age in month	0.042***	-0.054***	0.029	-0.032***
	(0.014)	(0.010)	(0.023)	(0.009)
BMI-for-age z-score (-1)	0.068**	0.001	-0.037	0.096***
	(0.033)	(0.029)	(0.074)	(0.028)
Birth order	-0.026	-0.026	0.113	0.037
	(0.038)	(0.033)	(0.073)	(0.031)
Child's highest grade	-0.093	0.536***	-0.192**	0.418***

	Maths score	Hours of study	Hours of economic work	Parental aspirations
	(0.102)	(0.047)	(0.098)	(0.041)
Non-Kinh child	-0.384**	-0.112	1.036***	0.143
	(0.159)	(0.137)	(0.285)	(0.131)
Number of children aged 0-5	-0.010	-0.073	-0.424***	-0.098
	(0.076)	(0.066)	(0.157)	(0.063)
Education of caregiver	0.045***	0.033***	-0.056**	0.064***
	(0.014)	(0.011)	(0.028)	(0.010)
Main caregiver is a parent	-0.136	0.146	0.872**	0.131
	(0.180)	(0.156)	(0.426)	(0.148)
Wealth index	0.195	0.582*	-0.845	3.006***
	(0.384)	(0.315)	(0.744)	(0.288)
North East	0.618***	-0.721***	2.953***	3.463***
	(0.170)	(0.135)	(0.347)	(0.140)
Red River Delta	0.145	0.361***	1.557***	3.639***
	(0.132)	(0.109)	(0.312)	(0.127)
South Central Coast	0.491***	-0.612***	1.867***	3.033***
	(0.147)	(0.115)	(0.314)	(0.123)
Mekong River Delta	0.435***	-0.588***	0.957***	4.500***

	Maths score	Hours of study	Hours of economic work	Parental aspirations
	(0.148)	(0.118)	(0.334)	(0.140)
Ln(average local wage)				5.815***
				(0.115)

Notes: Data source: 2006 and 2014 YLS; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A9. Probit models (marginal effects) for enrolment and schooling for age with the common sets of explanatory variables

	Enrolment			Schooling for age		
	All	Kinh	nonKinh	All	Kinh	nonKinh
Minority ethnic child	-0.005			-0.086***		
	(0.012)			(0.028)		
Age in month	-0.003**	-0.003**	-0.004	0.002	0.004*	-0.009**
	(0.001)	(0.001)	(0.004)	(0.002)	(0.002)	(0.004)
Younger cohort	-0.015	-0.043	0.008	0.378**	0.493***	-0.477*
	(0.097)	(0.097)	(0.385)	(0.187)	(0.186)	(0.281)
Boy	-0.019***	-0.021***	-0.002	-0.032***	-0.027**	-0.043
	(0.007)	(0.007)	(0.016)	(0.012)	(0.012)	(0.042)
Health problem	-0.015	0.025	-0.116*	-0.032	-0.011	-0.136
	(0.022)	(0.020)	(0.067)	(0.029)	(0.025)	(0.099)
Mother's year of schooling	0.005***	0.004***	0.005	0.006***	0.004**	0.029*
	(0.001)	(0.001)	(0.007)	(0.002)	(0.002)	(0.016)
Father's years of schooling	0.004***	0.003**	0.020***	0.006***	0.006***	0.008
	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.009)
Older siblings	-0.014***	-0.014***	-0.016	-0.015***	-0.014***	-0.023
	(0.004)	(0.004)	(0.012)	(0.005)	(0.005)	(0.018)

	Enrolment			Schooling for age		
	All	Kinh	nonKinh	All	Kinh	nonKinh
Asset index	0.243***	0.240***	0.348***	0.339***	0.294***	0.626***
	(0.024)	(0.024)	(0.051)	(0.035)	(0.042)	(0.100)
School travel time	-0.000	-0.000	-0.000	0.003***	0.005***	0.003*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
Pseudo R2	0.412	0.438	0.298	0.208	0.155	0.179
N	2,921	2,516	405	2,921	2,516	405

Notes: Data source: 2009 YLS; Significance level: * p<0.1; ** p<0.05; *** p<0.01; Standard errors in parentheses

Table A10. Multilevel Models for Maths and Vietnamese scores with the common sets of explanatory variables

	Maths			Vietnamese		
	All	Kinh	nonKinh	All	Kinh	nonKinh
Minority ethnicity child	-2.287***			-2.608***		
	(0.498)			(0.434)		
Age in month	0.024*	0.017	0.064***	0.024	0.035*	0.010
	(0.014)	(0.017)	(0.023)	(0.017)	(0.021)	(0.016)
Boy	-0.357**	-0.392**	0.024	-1.276***	-1.347***	-0.733**
	(0.154)	(0.178)	(0.289)	(0.162)	(0.177)	(0.340)
Health problem	-0.878***	-1.070***	0.266	-0.612***	-0.673***	-0.321
	(0.186)	(0.192)	(0.506)	(0.189)	(0.192)	(0.597)
Mother's years of school	0.069**	0.066*	0.042	0.068***	0.055**	0.057
	(0.032)	(0.034)	(0.075)	(0.025)	(0.024)	(0.074)
Unknown mother’s education	0.493	0.323	1.303*	0.435	0.128	1.316
	(0.343)	(0.391)	(0.677)	(0.356)	(0.326)	(1.100)
Father's years of school	0.054**	0.062**	0.036	0.094***	0.099***	0.131*
	(0.023)	(0.025)	(0.061)	(0.025)	(0.027)	(0.070)
Unknown farther’s education	0.599**	0.809***	-0.179	1.190***	1.449***	-0.066
	(0.258)	(0.291)	(0.657)	(0.314)	(0.318)	(0.708)

	Maths			Vietnamese		
	All	Kinh	nonKinh	All	Kinh	nonKinh
Older siblings	-0.133**	-0.185***	0.039	-0.242***	-0.247***	-0.254**
	(0.063)	(0.070)	(0.117)	(0.072)	(0.082)	(0.107)
Asset index	0.701	-0.391	4.418	1.230	0.535	5.470*
	(1.993)	(1.980)	(5.830)	(1.616)	(1.676)	(2.933)
School travel time	-0.015	-0.028*	0.000	0.002	0.000	0.005
	(0.013)	(0.015)	(0.019)	(0.010)	(0.012)	(0.018)
Constant	14.231***	16.412***	3.500	15.925***	15.541***	11.699***
	(2.467)	(2.728)	(4.753)	(2.270)	(2.606)	(3.579)
$\text{Ln}(\delta_v)$	0.947***	0.675***	1.586***	0.673***	0.489***	1.212***
	(0.156)	(0.152)	(0.122)	(0.146)	(0.142)	(0.106)
$\text{Ln}(\delta_u)$	0.904***	0.844***	0.754**	0.499***	0.484***	-10.408
	(0.127)	(0.120)	(0.319)	(0.106)	(0.100)	(52.502)
$\text{Ln}(\delta_\varepsilon)$	1.398***	1.408***	1.241***	1.372***	1.374***	1.297***
	(0.028)	(0.030)	(0.070)	(0.020)	(0.021)	(0.053)
N	3,218	2,831	387	3,218	2,831	387

Data source: 2011-12 YLSS; Significance level: * $p<0.1$; ** $p<0.05$; *** $p<0.01$; Standard errors in parentheses

Table A11. Multilevel mixed effect Probit models (marginal effects) for enrolment and schooling for age

	Enrolment	Schooling for age
Minority Ethnic child	0.003	-0.067*
	(0.013)	(0.037)
Age in month	-0.002*	0.002
	(0.001)	(0.002)
Younger cohort	0.026	0.371**
	(0.096)	(0.173)
Boy	-0.018***	-0.032**
	(0.006)	(0.012)
Height-for-age z score	0.003	0.027***
	(0.004)	(0.007)
Health problem	-0.002	-0.023
	(0.019)	(0.029)
Mother's year of schooling	0.005***	0.004*
	(0.001)	(0.002)
Father's years of schooling	0.005***	0.006***
	(0.001)	(0.002)
Older siblings	-0.012***	-0.009*
	(0.004)	(0.005)
Male head	-0.008	-0.026
	(0.014)	(0.021)
Asset index	0.224***	0.305***
	(0.031)	(0.036)
Health shock	-0.029***	-0.027**
	(0.008)	(0.013)
Newborn baby shock	-0.043**	-0.043*
	(0.017)	(0.024)
School travel time	-0.000	0.003***
	(0.000)	(0.001)
Paved road in commune	0.008	0.063***
	(0.012)	(0.021)
Factory in commune	0.016	-0.021

	Enrolment	Schooling for age
	(0.012)	(0.019)
Ln(population in commune)	-0.001	0.013
	(0.010)	(0.019)
N	2,921	2,921

Data source: 2009 YLS; Significance level: * p<0.1; ** p<0.05; *** p<0.01;
Standard errors in parentheses

Table A12. Probit models (marginal effects) and decomposition results for enrolment and schooling for age (without mother's education)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Age in month	-0.002*	-0.003	0.006	0.004**	-0.006	0.016
	(0.001)	(0.003)	(0.009)	(0.002)	(0.005)	(0.014)
Younger cohort	-0.001	0.112	0.002	0.502***	-0.223	-0.005
	(0.102)	(0.292)	(0.006)	(0.149)	(0.406)	(0.009)
Boy	-0.022***	0.001	0.000	-0.028**	-0.024	-0.000
	(0.008)	(0.034)	(0.000)	(0.012)	(0.042)	(0.001)
Height-for-age z score	0.004	0.006	-0.005	0.014**	0.118***	-0.099***
	(0.005)	(0.020)	(0.015)	(0.007)	(0.024)	(0.020)
Health problem	0.030	-0.078	-0.002	-0.008	-0.111*	-0.003
	(0.023)	(0.055)	(0.001)	(0.024)	(0.066)	(0.002)
Father's years of schooling	0.004***	0.022***	-0.077***	0.007***	0.007	-0.028
	(0.001)	(0.007)	(0.022)	(0.001)	(0.009)	(0.029)
Older siblings	-0.013***	-0.013	-0.006	-0.013**	-0.022	-0.011
	(0.003)	(0.011)	(0.005)	(0.006)	(0.014)	(0.007)
Male head	-0.010	-0.027	-0.001	-0.022	-0.053	-0.003
	(0.013)	(0.054)	(0.004)	(0.019)	(0.094)	(0.005)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Asset index	0.240***	0.301***	-0.054***	0.291***	0.388***	-0.079***
	(0.028)	(0.115)	(0.021)	(0.043)	(0.133)	(0.026)
Health shock	-0.026***	-0.038	0.000	-0.034***	0.015	-0.000
	(0.009)	(0.037)	(0.000)	(0.013)	(0.051)	(0.000)
Newborn baby shock	-0.027	-0.096**	-0.004*	-0.059**	-0.004	-0.000
	(0.021)	(0.047)	(0.002)	(0.024)	(0.071)	(0.003)
School travel time	-0.000	-0.000	-0.002	0.005***	0.003***	0.020***
	(0.000)	(0.001)	(0.004)	(0.001)	(0.001)	(0.007)
Paved road in commune	0.025	0.037	-0.007	0.063***	0.029	-0.006
	(0.016)	(0.049)	(0.010)	(0.023)	(0.064)	(0.013)
Factory in commune	0.005	0.055	-0.012	-0.028**	0.124**	-0.029**
	(0.009)	(0.049)	(0.011)	(0.014)	(0.063)	(0.015)
Ln(population in commune)	0.013	-0.043	0.025	0.033**	0.087*	-0.056*
	(0.009)	(0.043)	(0.025)	(0.013)	(0.048)	(0.031)
Total			-0.137***			-0.282***
			(0.023)			(0.028)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Pseudo R2	0.439	0.326		0.168	0.231	
N	2,516	405		2,516	405	

Notes: Data source: 2009 YLS; Significance level: * p<0.1; ** p<0.05; *** p<0.01; Standard errors in parentheses

Table A13. Probit models (marginal effects) and decomposition results for enrolment and schooling for age (without father's education)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Age in month	-0.002**	-0.002	0.004	0.004**	-0.007	0.016
	(0.001)	(0.004)	(0.011)	(0.002)	(0.005)	(0.012)
Younger cohort	-0.007	0.201	0.004	0.511***	-0.266	-0.005
	(0.099)	(0.299)	(0.007)	(0.150)	(0.403)	(0.008)
Boy	-0.022***	0.006	0.000	-0.027**	-0.024	-0.000
	(0.008)	(0.034)	(0.001)	(0.012)	(0.042)	(0.001)
Height-for-age z score	0.006	0.008	-0.007	0.015**	0.117***	-0.092***
	(0.005)	(0.020)	(0.016)	(0.007)	(0.024)	(0.019)
Health problem	0.026	-0.072	-0.002	-0.007	-0.115*	-0.003
	(0.023)	(0.054)	(0.002)	(0.024)	(0.066)	(0.002)
Mother's year of schooling	0.005***	0.008	-0.036	0.006***	0.022**	-0.089**
	(0.001)	(0.008)	(0.039)	(0.002)	(0.010)	(0.039)
Older siblings	-0.011***	-0.013	-0.006	-0.011*	-0.018	-0.009
	(0.003)	(0.011)	(0.006)	(0.006)	(0.015)	(0.007)
Male head	0.004	-0.020	-0.001	-0.006	-0.030	-0.002
	(0.012)	(0.057)	(0.004)	(0.018)	(0.098)	(0.005)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Asset index	0.248***	0.356***	-0.071***	0.317***	0.344***	-0.065***
	(0.026)	(0.117)	(0.024)	(0.042)	(0.133)	(0.025)
Health shock	-0.025***	-0.039	0.000	-0.035***	0.006	-0.000
	(0.009)	(0.038)	(0.000)	(0.013)	(0.052)	(0.000)
Newborn baby shock	-0.025	-0.111**	-0.005**	-0.060**	-0.014	-0.001
	(0.021)	(0.047)	(0.002)	(0.024)	(0.070)	(0.003)
School travel time	-0.000	-0.000	-0.003	0.005***	0.003***	0.019***
	(0.000)	(0.001)	(0.005)	(0.001)	(0.001)	(0.006)
Paved road in commune	0.021	0.031	-0.006	0.062***	0.020	-0.004
	(0.016)	(0.049)	(0.011)	(0.024)	(0.064)	(0.012)
Factory in commune	0.005	0.050	-0.012	-0.030**	0.115*	-0.025*
	(0.009)	(0.048)	(0.012)	(0.014)	(0.064)	(0.014)
Ln(population in commune)	0.015*	-0.035	0.022	0.035***	0.082*	-0.049*
	(0.009)	(0.043)	(0.028)	(0.013)	(0.050)	(0.029)
Total			-0.119***			-0.308***
			(0.031)			(0.027)

	Enrolment			SAGE		
	Kinh	nonKinh	Explained	Kinh	nonKinh	Explained
Pseudo R2	0.445	0.308		0.164	0.237	
N	2,516	405		2,516	405	

Data source: 2009 YLS; Significance level: * p<0.1; ** p<0.05; *** p<0.01; Standard errors in parentheses

Table A14. Mean of some explanatory variables by parents' education

	Father's education is known	Father's education is unknown	Difference
Minor Ethnicity	0.132	0.079	0.053***
Age in month	124.985	125.669	-0.685***
Boy	0.531	0.503	0.028
Health problem	0.224	0.218	0.006
Repeated grades	0.038	0.050	-0.012
Vietnamese at home	0.901	0.951	-0.050***
Older siblings	0.973	0.886	0.087*
Asset index	0.715	0.714	0.002
School travel time	0.130	0.119	0.012
	Mother's education is known	Mother's education is unknown	Difference
Minor Ethnicity	0.130	0.080	0.050***
Age in month	124.912	126.135	-1.222***
Boy	0.523	0.532	-0.009
Health problem	0.222	0.223	-0.000
Repeated grades	0.039	0.051	-0.013
Vietnamese at home	0.901	0.960	-0.059***
Older siblings	0.970	0.880	0.089*
Asset index	0.715	0.716	-0.002
School travel time	0.129	0.121	0.008

Data source: 2011-12 YLSS; Significance level: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table A15. Benchmark estimations

Victimisation rate	(1)	(2)	(3)	(4)
Lagged victim rate	0.117***	0.122***	0.110***	0.113***
	(0.033)	(0.033)	(0.033)	(0.032)
Years of school, aged 15+	**-0.854****	**-0.853*****	**-0.929****	**-0.932*****
	(0.335)	**(0.330)**	**(0.365)**	**(0.352)**
Year = 2009	-1.326***	-1.344***	-1.310***	-1.329***
	(0.116)	(0.115)	(0.117)	(0.115)
Year = 2010	-0.932***	-0.979***	-0.989***	-1.045***
	(0.226)	(0.216)	(0.224)	(0.210)
Year = 2011	-1.177***	-1.245***	-1.461***	-1.549***
	(0.167)	(0.158)	(0.155)	(0.151)
Year = 2012	-1.244***	-1.330***	-1.606***	-1.714***
	(0.216)	(0.203)	(0.195)	(0.184)
Ln(Population)		0.944***		1.127***
		(0.360)		(0.371)
Urban population (%)		0.005		0.004
		(0.005)		(0.005)
Male, aged 15-25 (%)		-0.009		-0.010
		(0.019)		(0.019)

Victimisation rate	(1)	(2)	(3)	(4)
Ln(per capita expenditure)			0.770**	0.788**
			(0.366)	(0.368)
Gini			1.791**	1.854**
			(0.873)	(0.869)
Poverty (%)			0.001	0.001
			(0.002)	(0.002)
GDP growth (%)			0.002	0.002
			(0.005)	(0.005)
Number of IVs	10	13	14	17
Sargan Chi_sq	2.442	2.969	2.072	2.710
Sargan P_value	0.655	0.563	0.723	0.607
Hasen Chi_sq	3.097	3.709	2.040	2.744
Hasen P_value	0.542	0.447	0.728	0.602
AR(1) z	0.192	0.287	0.195	0.277
AR(1) P_value	0.848	0.774	0.845	0.782
CLR	0.031	0.04	0.022	0.029
CLR P_value	0.987	0.984	0.991	0.988
AR	2.927	3.467	1.945	1.954
AR P_value	0.818	0.748	0.925	0.924
Number of group	507	507	507	507

Victimisation rate	(1)	(2)	(3)	(4)
Number of observation	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A16. Benchmark estimations, controlling for the share of non-active population

	(1)	(2)
Lagged victim rate	0.111***	0.115***
	(0.032)	(0.032)
Year of school, aged 15+	**-0.898****	**-0.900*****
	(0.356)	**(0.345)**
Year = 2009	-1.323***	-1.342***
	(0.118)	(0.116)
Year = 2010	-1.009***	-1.063***
	(0.221)	(0.207)
Year = 2011	-1.479***	-1.563***
	(0.158)	(0.153)
Year = 2012	-1.630***	-1.735***
	(0.196)	(0.186)
Ln(population)		1.106***
		(0.366)
Urban population (%)		0.004
		(0.005)
Male, aged 15-25 (%)		-0.010
		(0.019)
Ln(per capita expenditure)	0.743**	0.759**
	(0.362)	(0.364)
Gini	1.848**	1.909**
	(0.873)	(0.870)
Poverty (%)	0.001	0.001
	(0.002)	(0.002)
GDP growth (%)	0.002	0.002
	(0.005)	(0.005)
Non-active population, aged 15+ (%)	**-2.680**	**-2.628**
	(3.965)	**(3.993)**
Number of instruments	15	18
Sargan Chi_sq	2.123	2.766
Sargan P_value	0.713	0.598
Hasen Chi_sq	2.223	2.970
Hasen P_value	0.695	0.563

	(1)	(2)
AR(1) z	0.295	0.375
AR(1) P_value	0.768	0.707
CLR	0.026	0.082
CLR P_value	0.989	0.968
AR	2.124	2.186
AR P_value	0.908	0.902
Number of districts	507	507
Number of observations	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A17. Robustness check

Victimisation rate	(1) Lag 2 only	(2) Lag 2-3	(3) Lag 2-4	(4) Lag 2-5	(5) Collapsed lags 2-4
Lagged victim rate	0.092***	0.085***	0.086***	0.089***	0.093***
	(0.024)	(0.024)	(0.024)	(0.025)	(0.032)
Years of school, aged 15+	**-0.866*** **	**-0.909*** **	**-0.932*** **	**-0.866*** **	**-0.929*** **
	(0.200)	**(0.198)**	**(0.200)**	**(0.196)**	**(0.253)**
Year = 2009	-1.322***	-1.300***	-1.293***	-1.300***	-1.410***
	(0.099)	(0.099)	(0.099)	(0.101)	(0.135)
Year = 2010	-1.106***	-1.082***	-1.066***	-1.084***	-1.333***
	(0.152)	(0.150)	(0.151)	(0.155)	(0.311)
Year = 2011	-1.609***	-1.604***	-1.600***	-1.581***	-1.736***
	(0.137)	(0.139)	(0.139)	(0.143)	(0.502)
Year = 2012	-1.776***	-1.783***	-1.777***	-1.774***	-1.945***
	(0.158)	(0.158)	(0.158)	(0.164)	(0.677)
Ln(Population)	1.130***	1.313***	1.307***	1.262***	1.100
	(0.327)	(0.336)	(0.329)	(0.352)	(0.695)
Urban population (%)	0.002	0.005	0.005	0.006	0.002
	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)
Male, aged 15-25 (%)	-0.017	-0.016	-0.015	-0.016	-0.014

Victimisation rate	(1) Lag 2 only	(2) Lag 2-3	(3) Lag 2-4	(4) Lag 2-5	(5) Collapsed lags 2-4
	(0.018)	(0.019)	(0.019)	(0.019)	(0.020)
Ln(per capita expenditure)	0.822***	0.921***	0.959***	0.912***	2.741
	(0.298)	(0.307)	(0.295)	(0.296)	(2.323)
Gini	1.731**	1.703*	1.713**	1.442*	-8.231
	(0.835)	(0.871)	(0.859)	(0.875)	(5.755)
Poverty (%)	0.001	0.001	0.001	0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
GDP growth (%)	0.002	0.002	0.001	0.001	0.063
	(0.005)	(0.005)	(0.005)	(0.005)	(0.044)
Number of IVs	19	25	29	31	25
Sargan Chi_sq	5.135	15.315	17.091	18.228	6.835
Sargan P_value	0.527	0.225	0.380	0.441	0.868
Hasen Chi_sq	5.014	15.384	16.233	22.800	8.868
Hasen P_value	0.542	0.221	0.437	0.198	0.714
AR(1) Z_value	0.232	0.203	0.184	0.272	-0.407
AR(1) P_value	0.817	0.839	0.854	0.786	0.684
CLR	0.012	0.26	0.265	0.422	3.695
CLR P_value	0.995	0.901	0.905	0.857	0.995
AR	4.884	14.668	15.527	22.193	10.274

Victimisation rate	(1)	(2)	(3)	(4)	(5)
	Lag 2 only	Lag 2-3	Lag 2-4	Lag 2-5	Collapsed lags 2-4
AR P_value	0.77	0.401	0.625	0.33	0.923
Number of group	507	507	507	507	507
Number of observation	1,812	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A18. Different measurements of education

Victimisation rate	(1)	(2)	(3)
Lagged victim rate	0.122***	0.130***	0.085**
	(0.032)	(0.034)	(0.039)
Year of school, aged 7+	**-1.074*** **		
	(0.417)		
No degree, aged 15+ (%)		**0.122*** **	
		(0.040)	
Primary education, aged 15+ (%)			**0.066**
			(0.082)
Junior Secondary education, aged 15+ (%)			**-0.134***
			(0.071)
Senior Secondary education and above, aged 15+ (%)			**-0.172*** **
			(0.062)
Year = 2009	-1.259***	-1.419***	-1.604***
	(0.134)	(0.102)	(0.151)
Year = 2010	-1.043***	-1.181***	-1.545***
	(0.212)	(0.164)	(0.237)
Year = 2011	-1.519***	-1.576***	-1.987***
	(0.155)	(0.151)	(0.224)
Year = 2012	-1.681***	-1.714***	-2.096***
	(0.191)	(0.180)	(0.257)
Ln(population)	1.141***	0.899***	1.028**
	(0.376)	(0.312)	(0.422)
Urban population (%)	0.004	0.014**	0.036***
	(0.005)	(0.006)	(0.013)
Male, aged 15-25 (%)	0.002	0.007	0.012
	(0.020)	(0.020)	(0.026)
Ln(per capita expenditure)	0.785**	0.650**	1.374***
	(0.374)	(0.318)	(0.394)
Gini	1.765**	1.160	2.517**
	(0.868)	(0.904)	(1.233)
Poverty (%)	0.002	0.001	0.001

Victimisation rate	(1)	(2)	(3)
	(0.002)	(0.002)	(0.003)
GDP growth (%)	0.001	0.003	-0.004
	(0.005)	(0.005)	(0.006)
Number of instruments	17	17	23
Sargan test: Chi squared	3.220	4.216	6.269
Sargan test: P value	0.522	0.378	0.617
Hansen test: Chi squared	3.166	7.147	8.712
Hasen test: P value	0.531	0.128	0.367
Test of AR(1): Z	0.507	-0.181	-0.103
Test of AR(1): P value	0.612	0.856	0.918
Test of weak IV: CLR	0.057	0.861	1.215
Test of weak IV: CLR P value	0.976	0.693	0.906
Test of weak IV: AR	2.965	6.457	7.931
Test of weak IV: AR P value	0.813	0.374	0.79
Number of districts	507	507	507
Number of observations	1,812	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A19. School age children

Victimisation rate	All population	Aged 7-18
Lagged victim rate	0.121***	0.082***
	(0.028)	(0.024)
Enrolment rate, age7-18	**-0.088****	**0.012**
	(0.044)	**(0.041)**
Year = 2009	-1.439***	-1.574***
	(0.098)	(0.095)
Year = 2010	-1.411***	-1.638***
	(0.117)	(0.118)
Year = 2011	-1.552***	-1.641***
	(0.149)	(0.153)
Year = 2012	-1.757***	-1.835***
	(0.175)	(0.185)
Ln(population)	0.822**	0.418
	(0.364)	(0.357)
Urban population (%)	0.003	0.000
	(0.004)	(0.005)
Male, aged 15-25 (%)	-0.046*	0.003
	(0.025)	(0.026)
Ln(per capita expenditure)	0.370	0.176
	(0.297)	(0.252)
Gini	1.625*	0.947
	(0.872)	(0.769)
Poverty (%)	-0.000	0.002***
	(0.001)	(0.001)
GDP growth (%)	0.001	0.008*
	(0.005)	(0.005)
Number of instruments	17	17
Sargan test: Chi squared	4.542	3.273
Sargan test: P value	0.338	0.513
Hansen test: Chi squared	7.594	5.736
Hasen test: P value	0.108	0.220
Test of AR(1): Z	1.320	1.134

Victimisation rate	All population	Aged 7-18
Test of AR(1): P value	0.187	0.257
Test of weak IV: CLR	0.163	0.548
Test of weak IV: CLR P value	0.95	0.872
Test of weak IV: AR	7.19	6.342
Test of weak IV: AR P value	0.304	0.386
Number of districts	507	507
Number of observations	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A20. Gender difference

Victimisation rate	All population	Female	All population	Male
Lagged victim rate	0.109***	0.099***	0.117***	0.125***
	(0.033)	(0.027)	(0.032)	(0.034)
Years of schooling, female aged 15+	**-0.834*** **	**-0.499***		
	(0.314)	**(0.299)**		
Years of schooling, male aged 15+			**-1.011** **	**-1.383** **
			(0.413)	**(0.547)**
Year = 2009	-1.345***	-1.502***	-1.321***	-1.170***
	(0.112)	(0.110)	(0.120)	(0.139)
Year = 2010	-1.068***	-1.378***	-1.051***	-0.773***
	(0.205)	(0.194)	(0.219)	(0.288)
Year = 2011	-1.536***	-1.648***	-1.581***	-1.495***
	(0.154)	(0.138)	(0.149)	(0.180)
Year = 2012	-1.691***	-1.764***	-1.762***	-1.721***
	(0.191)	(0.171)	(0.180)	(0.225)
Ln(population)	1.060***	0.863**	1.172***	1.435**
	(0.350)	(0.353)	(0.401)	(0.564)

Victimisation rate	All population	Female	All population	Male
Urban population (%)	0.003	0.001	0.004	0.008
	(0.005)	(0.004)	(0.005)	(0.007)
Male, aged 15-25 (%)	-0.006	-0.017	-0.016	-0.013
	(0.020)	(0.016)	(0.020)	(0.026)
Ln(per capita expenditure)	0.650*	0.375	0.926**	1.241**
	(0.338)	(0.336)	(0.416)	(0.506)
Gini	1.859**	1.326*	1.871**	2.156**
	(0.873)	(0.801)	(0.880)	(1.094)
Poverty (%)	0.001	0.001	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
GDP growth (%)	0.004	0.005	0.000	-0.002
	(0.005)	(0.004)	(0.005)	(0.006)
Number of instruments	17	17	17	17
Sargan test: Chi squared	2.375	4.035	2.731	1.931
Sargan test: P value	0.667	0.401	0.604	0.749
Hansen test: Chi squared	2.060	3.788	3.305	1.531
Hasen test: P value	0.725	0.435	0.508	0.821
Test of AR(1): Z	0.144	0.635	0.507	0.152

Victimisation rate	All population	Female	All population	Male
Test of AR(1): P value	0.885	0.525	0.612	0.879
Test of weak IV: CLR	0.042	0.013	0.097	0.052
Test of weak IV: CLR P value	0.982	0.997	0.965	0.978
Test of weak IV: AR	2.037	3.64	3.07	1.571
Test of weak IV: AR P value	0.916	0.725	0.8	0.955
Number of districts	507	507	507	507
Number of observations	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses

Table A21. Interaction with poverty

	5th percentile	10th percentile	Poverty line	15th percentile	20th percentile
Lagged victim rate	0.085***	0.098***	0.116***	0.110***	0.128***
	(0.032)	(0.031)	(0.032)	(0.034)	(0.041)
Year of school, aged 15+	**-1.247*** **	**-1.114*** **	**-0.988*** **	**-1.027*** **	**-1.183*** **
	(0.452)	**(0.384)**	**(0.369)**	**(0.357)**	**(0.425)**
Interaction	**0.015***	**0.007***	**0.002**	**0.004**	**0.008**
	(0.008)	**(0.004)**	**(0.002)**	**(0.003)**	**(0.007)**
Year = 2009	-1.354***	-1.347***	-1.322***	-1.331***	-1.261***
	(0.114)	(0.110)	(0.114)	(0.111)	(0.137)
Year = 2010	-1.273***	-1.255***	-1.032***	-1.184***	-1.002***
	(0.196)	(0.177)	(0.203)	(0.167)	(0.202)
Year = 2011	-1.900***	-1.845***	-1.531***	-1.735***	-1.632***
	(0.209)	(0.192)	(0.147)	(0.179)	(0.166)
Year = 2012	-2.202***	-2.139***	-1.689***	-1.995***	-1.856***
	(0.265)	(0.246)	(0.176)	(0.227)	(0.207)
Ln(population)	1.127***	1.125***	1.090***	1.077***	1.236***
	(0.400)	(0.379)	(0.379)	(0.374)	(0.438)
Urban population (%)	0.004	0.004	0.004	0.004	0.005

	5th percentile	10th percentile	Poverty line	15th percentile	20th percentile
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Male, aged 15-25 (%)	0.004	-0.004	-0.011	-0.004	-0.004
	(0.025)	(0.023)	(0.020)	(0.022)	(0.020)
Ln(per capita expenditure)	3.155**	2.581**	0.782**	1.975*	1.707*
	(1.477)	(1.215)	(0.388)	(1.088)	(0.921)
Gini	-3.781	-2.193	1.716*	-0.742	0.968
	(3.384)	(2.626)	(0.892)	(2.245)	(1.326)
Poverty (%)	0.002	0.002	-0.019	0.002	0.002
	(0.002)	(0.002)	(0.018)	(0.002)	(0.002)
GDP growth (%)	0.000	0.002	0.002	0.002	0.002
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Number of instruments	20	20	20	20	20
Sargan test: Chi squared	1.182	0.896	2.487	0.837	2.087
Sargan test: P value	0.978	0.989	0.870	0.991	0.911
Hansen test: Chi squared	1.856	1.715	3.501	2.163	4.317
Hasen test: P value	0.932	0.944	0.744	0.904	0.634
Test of AR(1): Z	-0.504	-0.099	0.255	0.276	0.222
Test of AR(1): P value	0.615	0.921	0.799	0.783	0.824

	5th percentile	10th percentile	Poverty line	15th percentile	20th percentile
Test of weak IV: CLR	1.443	0.916	0.36	4.526	0.673
Test of weak IV: CLR P value	0.772	0.866	0.97	0.29	0.921
Test of weak IV: AR	3.253	2.663	3.675	6.581	4.284
Test of weak IV: AR P value	0.953	0.976	0.931	0.681	0.892
Number of districts	507	507	507	507	507
Number of observations	1,812	1,812	1,812	1,812	1,812

Notes: Data sources: Susenas, DAPOER, MoF; Significance level: *10%; **5%; ***1%; Standard errors are in parentheses