EDUCATION ECONOMICS, 2018
https://doi.org/10.1080/09645292.2018.1444147

Routledge
Taylor & Francis Group

Ethnic gaps in child education outcomes in Vietnam: an investigation using Young Lives data

Hieu T. M. Nguyen

Arndt-Corden Department of Economics, Crawford School of Public Policy, Australian National University, Canberra, Australia

ABSTRACT

There are large gaps in child education outcomes between the Kinh majority and non-Kinh minorities in Vietnam. This paper seeks to understand the reasons for these ethnic gaps. The examination employs Probit and multilevel regression models, and associated decomposition techniques. The results show that Vietnam's ethnic gap in school enrolment is mostly attributable to household characteristics such as household expenditure and father's education. Gaps in schooling progress and performance are explained by a broader set of variables such as child, household, commune, school, and peer characteristics.

ARTICLE HISTORY
Received 30 March 2017
Accepted 16 February 2018

KEYWORDS
Ethnic gap; education outcomes; decomposition; Vietnam

JEL
I24; O15

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% of the population; the share of all other ethnic groups is 14%. The largest minority ethnic groups, such as Tay, Thai, Muong and Khmer, account for less than 2%. The living standards of minority groups are much lower than those of the Kinh. For example, the poverty rate in 2010 was 12.9% for the Kinh while more than 66% 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 2010a).

Ethnic inequality in education is also considerable. The results of the 2009 Vietnam population and housing census (GSO 2010a) show that the literacy rate of the population aged 10 and over is 96% for Kinh compared to 78% for minority groups. 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%, 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% compared to 3% 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 9–20 (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

CONTACT Hieu T. M. Nguyen nguyen.hieu@anu.edu.au; nguyen.t.minh.hieu@gmail.com
Supplemental data for this article can be accessed at https://doi.org/10.1080/09645292.2018.1444147.

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 my knowledge multilevel model decomposition techniques have not yet been used in the literature to examine education outcomes.

The rest of this paper is structured as follows. Section 2 introduces the two data sets used in the analysis. Section 3 presents methodologies applied in the paper. The estimation and decomposition results for different education outcomes are discussed in Section 4. Section 5 concludes.

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 1 displays summary statistics concerning education and ethnicity calculated from YLS and two other national representative samples, the 15% sample of 2009 Vietnam Population and Housing Census (Minnesota Population Center 2017), and 2010 Vietnam Household Living Standard Survey (GSO 2010b). 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% lower.

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%, almost the same as that of the older cohort in the 2009 YLS (see Table 1).

Table 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 year old)	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 year old)	76.2	76.7	80.1
Kinh	80.0	81.7	83.6
Non-Kinh	50.6	62.9	69.2

Source: Author's calculation from 2009 Young Lives Survey, 15% sample of Vietnam Population and Housing Census, 2010 Vietnam Household Living Standard Survey.

Table 2. Sampling of 2009YLS and 2011–2012YLSS.

	Ethnicity		Gender		
	Kinh	Non-Kinh	Boys	Girls	Total
YLS	2516	405	1483	1438	2921
Younger cohort (7–8 year old)	1671	279	1004	946	1950
Older cohort (14–15 year old)	845	126	481	490	971
YLSS	2879	399	1734	1550	3284

Data source: 2009 YLS and 2011–2012 YLSS.

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 2). The share of minority ethnic children is 13.8% in the YLS.

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%, 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 2).

3. Methodology

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.

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

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; v_{jk} are class level error terms, representing class random effects; and ε_{ijk} are individual level error terms.

The estimated equation (1) is

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

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.

$$\Delta^{\text{multilevel}} = \bar{S}^M - \bar{S}^m \tag{3}$$

$$\Delta^{\text{multilevel}} = \sum_{i=1}^{3} \hat{\beta}_i(\bar{X}_i^M - \bar{X}_i^m) + (\bar{\hat{v}}^M - \bar{\hat{v}}^m) + (\bar{\hat{u}}^M - \bar{\hat{u}}^m) - \hat{\gamma} \tag{4}$$

In equation (4), $\sum_{i=1}^{3} \hat{\beta}_i(\bar{X}_i^M - \bar{X}_i^m)$ is the explained part, providing the contribution of explanatory variables to the ethnic gap in the test scores and $(\bar{\hat{v}}^M - \bar{\hat{v}}^m) + (\bar{\hat{u}}^M - \bar{\hat{u}}^m)$ is the random part generated by school and class level errors; $\hat{\gamma}$ is the unexplained component created by unobservable factors.

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:

$$P(Y = 1|X) = F(X\beta) \tag{5}$$

where $Y = 1$ if the child enrols in school, and $Y = 0$ otherwise. The right-hand side of equation (5) is the conditional probability that $Y = 1$. *F* is a cumulative normal distribution function. Equation (5) can be estimated by maximum likelihood methods.

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

$$\Delta^{\text{Probit}} = \overline{P}^M - \overline{P}^m = \overline{F(X^M\beta^M)} - \overline{F(X^m\beta^m)} \tag{6}$$

$$\Delta^{\text{Probit}} = \{\overline{F(X^M\beta^M)} - \overline{F(X^m\beta^M)}\} + \{\overline{F(X^m\beta^M)} - \overline{F(X^m\beta^m)}\} \tag{7}$$

The first component in the right-hand side of equation (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.

3.3. Variable selection

The definition and descriptive statistics of variables used in the empirical analysis are shown in Tables 3 and 4.

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.

The enrolment rate of the younger cohort in the YLS is 98.5% and that of the older cohort is 76.2%. The ethnic gap in the enrolment rate for the former group of children is more than 8% and that for the latter is almost 30%. For all Young Lives children, the ethnic gap in the enrolment rate is 14.4% (see Table 4).

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

$$\text{SAGE} = \frac{\text{the highest grade attainted by the child}}{\text{the child's age} - 6} \tag{8}$$

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 1). If a child starts school at age 6, does not repeat any grade and continues to enrol in

Table 3. Definition of variables in the models.

Variables	Definition
From the YLS and YLSS	
Minority ethnicity child	Dummy variable: 1 if non-Kinh ethnicity; 0 if Kinh ethnicity
Age in months	Age in month of the child
Boy	Dummy variable: 1 for boy; 0 for girl
Health problem	Dummy variable: 1 if any long term health problems except vision related problems; 0 no health problems
Father's years of schooling	Father's years of schooling
Mother's years of schooling	Mother's years of schooling
Older siblings	Number of older siblings
School travel time	Minutes to travel from home to school
Asset index	Index of assets owned by the household
From the YLS only	
School enrolment	Dummy variable: 1 if currently enrol in school; 0 if not currently enrol
Schooling for age index	SAGE = Highest grade attainted/(age − 6)
Schooling for age index = 1	Dummy variable: 1 if SAGE = 1; 0 if SAGE < 1
Younger cohort	Dummy variable: 1 if born in 2000/2001; 0 if born in 1994/1995
Height-for-age *z* score	Height-for-age *z* score
Male head	Dummy variable: 1 if the household head is male; 0 if otherwise
Health shock	Dummy variable: 1 if experience of serious illness of a household members; 0 if otherwise
Newborn baby shock	Dummy variable: 1 if a newborn baby in the family since the last survey; 0 if otherwise
Paved road in commune	Dummy variable: 1 if there is a paved road in the commune; 0 if otherwise
Factory in commune	Dummy variable: 1 if there is a factory in the commune; 0 if otherwise
Ln(population in commune)	Natural logarithm of the commune's population
From YLSS only	
Raw maths scores	Number of correct answer in the maths test
IRT-adjusted maths scores	Maths scores adjusted by IRT
Raw Vietnamese scores	Number of correct answer in the Vietnamese test
IRT-adjusted Vietnamese scores	Vietnamese scores adjusted by IRT
Speaking Vietnamese at home	Dummy variable: 1 if the child speaks Vietnamese at home; 0 if otherwise
School travel time	Time to travel to school in minutes
Unknown mother's education	Dummy variable: 1 if mother's highest education is unknown by the child, 0 if otherwise
Unknown father's education	Dummy variable: 1 if father's highest education is unknown by the child, 0 if otherwise
Days absent from school of classmates	Average number of days absent from school of the child's classmates
Grade repetition of classmates	Number of the child's classmates repeated grades
Class size	Number of students in the child's class
Television in classroom	Dummy variable: 1 if there is a television in the classroom; 0 if otherwise
Dropout at school	Dummy variable: 1 if the school has dropout problem
School accesses to internet	Dummy variable: 1 if the school access to internet; 0 if otherwise
Newly established school	Dummy variable: 1 if the school were established less than 5 year ago; 0 if otherwise

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% of Kinh children started school at 6 or earlier compared to 75% 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% and that of non-Kinh children is 59% (see Table 4).

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

Table 4. 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***
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***

Data source: 2009 YLS and 2011–2012 YLSS.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

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 2 and Figure 3. The gaps exist in both maths and Vietnamese test scores and range from 4.4 points to 4.6 points (see Table 4).

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 Young Lives (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.[1]

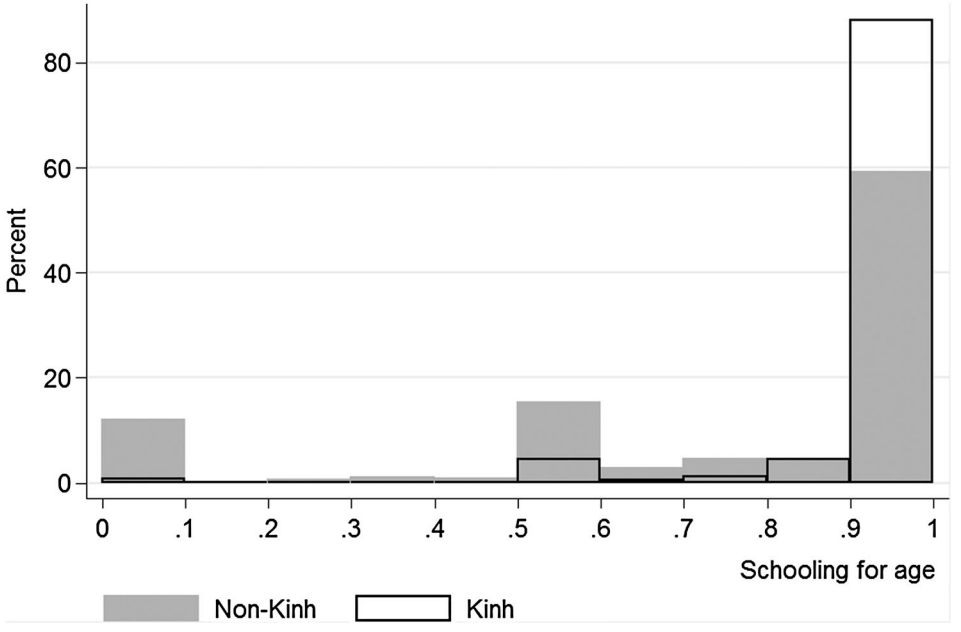

Figure 1. Histogram of schooling for age index by ethnic groups.

In the school survey, most of the information about household characteristics was collected by interviewing students. One problem is that 20% 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 parents' education is known by the child, and a dummy for 'Don't know' answers. Peer effects are

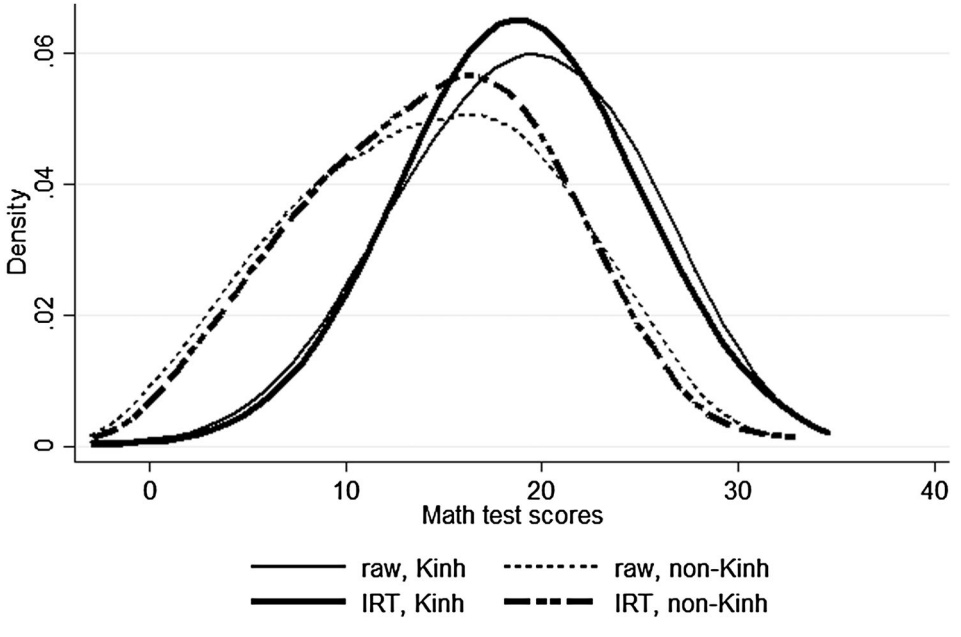

Figure 2. Kernel density curves of maths test scores by ethnic groups: IRT-adjusted and raw.

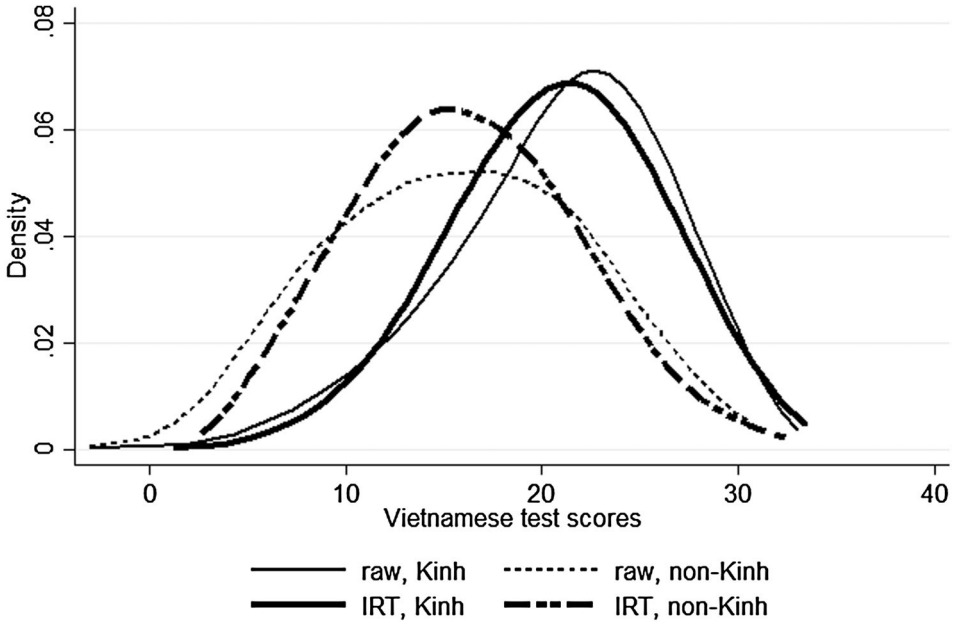

Figure 3. Kernel density curves of Vietnames test scores by ethnic groups: IRT-adjusted and raw.

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. Empirical results[2]

4.1. Enrolment

The Probit estimation for school enrolment[3] and the corresponding decomposition results are presented in Table 5.

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 absence of the ethnic gap but means that the child's ethnicity does not affect 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 influence 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 reduce the enrolment rate. Additionally, the number of older siblings significantly reduces the chance that a child participates 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.

4.1.2. Decomposition results

The decomposition result shows that 99% 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

Table 5. 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)	
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)	
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 R^2	0.425	0.450	0.327		
N	2921	2516	405		

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

children. The dominant role of father's education, which will be further explained in section 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% 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 4.3.2. 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% and household characteristics contribute to the total ethnic enrolment gap of 105%. 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.

4.2. Schooling for age (SAGE)

Recall that the distribution of SAGE piles up at 1 (see Figure 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 6.

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,

Table 6. 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
	(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
	(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 R^2	0.223	0.171	0.238		
N	2921	2516	405		

Data source: 2009 YLS.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Standard errors in parentheses.

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.

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.[4] 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 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 (Nørlund, Gates, and Vu 1995; Nguyen as cited by Nguyen et al. 2012), e.g. enrolment decisions, in the family, especially in a low income family (Dang and Le as cited by 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 (Nguyen et al. 2012).[5]

4.2.2. Decomposition results

The total explained part accounts for 106.5% of the gap in SAGE.[6] Over 100% of the ethnic gap explained in the model implies that unobservable factors help to reduce the gap by 6.5%. 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% 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 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% of the ethnic gap in enrolment but only under 6% in SAGE. In contrast, the explanatory power of mother's education in the latter is more than double that in the former.[7]

4.3. Schooling performance

The estimation and decomposition results for multilevel models of maths and Vietnamese test scores are presented in Tables 7 and 8, respectively.

Table 7. Multilevel models and decomposition results for maths test scores.

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	%Explained
Minority ethnic child	−1.739***				
	(0.557)				
Age in month	0.026*	0.018	0.074***	−0.041	−0.9
	(0.015)	(0.017)	(0.024)	(0.027)	
Boy	−0.353**	−0.373**	0.030	0.005	0.1
	(0.154)	(0.179)	(0.302)	(0.011)	
Health problem	−0.852***	−1.019***	0.429	−0.010	−0.2
	(0.187)	(0.193)	(0.504)	(0.019)	
Speaking Vietnamese at home	1.084**	2.143***	0.228	1.098***	24.6
	(0.439)	(0.693)	(0.454)	(0.249)	
Mother's years of school	0.066**	0.065*	0.040	0.343***	7.7
	(0.032)	(0.034)	(0.071)	(0.149)	
Unknown mother's education	0.420	0.266	1.296*	0.034	0.8
	(0.345)	(0.394)	(0.687)	(0.027)	
Father's years of school	0.050**	0.060**	0.017	0.152***	3.4
	(0.023)	(0.025)	(0.063)	(0.074)	
Unknown father's education	0.569**	0.789***	−0.383	0.054***	1.2
	(0.259)	(0.295)	(0.619)	(0.027)	
Older siblings	−0.125**	−0.175**	0.034	0.064***	1.4
	(0.063)	(0.070)	(0.120)	(0.031)	
Asset index	0.426	−0.324	9.088***	0.040	0.9
	(2.005)	(2.007)	(3.135)	(0.124)	
School travel time	−0.014	−0.028*	0.002	0.086*	1.9
	(0.013)	(0.015)	(0.019)	(0.065)	
Days absent from school of classmates	−0.685	0.212	−4.489***	0.132	3.0
	(0.566)	(0.551)	(1.013)	(0.106)	
Repeated grades of classmates	−12.364***	−10.838***	−19.672***	0.417***	9.3
	(4.181)	(4.110)	(5.309)	(0.144)	
Class size	0.000	−0.003	−0.273***	0.249	5.6
	(0.037)	(0.032)	(0.106)	(0.440)	
Television in classroom	1.882**	2.319**	12.034***	0.172***	3.8
	(0.897)	(0.903)	(2.579)	(0.086)	
Dropout at school	−2.072	−2.260*	−1.108	0.339*	7.6
	(1.620)	(1.349)	(3.142)	(0.263)	
School accesses to internet	0.569	0.035	5.511***	0.156	3.5
	(0.733)	(0.699)	(1.598)	(0.215)	
Newly established school	0.751	−0.782	3.933***	−0.034	−0.8
	(1.893)	(1.620)	(1.041)	(0.111)	
Constant	13.845***	14.665***	5.747		
	(2.593)	(2.924)	(4.061)		
Total				3.258	73.0
Mean predicted Maths score	14.470	18.766	14.123		
$Ln(\delta_v)$	0.762***	0.480**	1.336***		
	(0.218)	(0.216)	(0.160)		
$Ln(\delta_u)$	0.878***	0.820***	−20.218		
	(0.128)	(0.119)	(33.649)		
$Ln(\delta_\varepsilon)$	1.397***	1.406***	1.248***		
	(0.028)	(0.029)	(0.067)		
N	3218	2831	387		

Data source: 2010–2011 YLSS.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Standard errors in parentheses; Standard errors in the decomposition results are bootstrapped with 1000 replications.

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.[8]

Table 8. Multilevel models and decomposition results for Vietnamese test scores.

	Estimation			Decomposition	
	All	Kinh	Non-Kinh	Explained	%Explained
Minority ethnic child	−1.613***				
	(0.463)				
Age in month	0.026	0.035*	0.016	−0.040	−0.9
	(0.017)	(0.021)	(0.016)	(0.032)	
Boy	−1.268***	−1.330***	−0.698**	0.019	0.4
	(0.163)	(0.180)	(0.340)	(0.035)	
Health problem	−0.579***	−0.630***	−0.263	−0.007	−0.1
	(0.187)	(0.188)	(0.594)	(0.014)	
Speaking Vietnamese at home	2.009***	2.381***	1.487**	1.649***	35.3
	(0.470)	(0.648)	(0.681)	(0.293)	
Mother's years of school	0.063**	0.054**	0.031	0.328***	7.0
	(0.025)	(0.024)	(0.073)	(0.113)	
Unknown mother's education	0.323	0.074	1.115	0.027	0.6
	(0.341)	(0.325)	(1.050)	(0.026)	
Father's years of school	0.087***	0.097***	0.093	0.265***	5.7
	(0.025)	(0.027)	(0.076)	(0.080)	
Unknown father's education	1.135***	1.433***	−0.342	0.104***	2.2
	(0.312)	(0.319)	(0.748)	(0.037)	
Older siblings	−0.233***	−0.238***	−0.254**	0.115***	2.5
	(0.072)	(0.082)	(0.113)	(0.038)	
Asset index	0.836	0.418	5.877**	0.063	1.4
	(1.667)	(1.747)	(2.995)	(0.103)	
School travel time	0.004	0.000	0.011	−0.005	−0.1
	(0.010)	(0.012)	(0.018)	(0.051)	
Days absent from school of classmates	−0.676	−0.185	−2.390**	0.131*	2.8
	(0.504)	(0.570)	(1.103)	(0.092)	
Repeated grades of classmates	−5.372*	−4.459	−3.498	0.193**	4.1
	(3.039)	(3.139)	(4.397)	(0.100)	
Class size	0.015	0.009	−0.035	0.396	8.5
	(0.029)	(0.030)	(0.077)	(0.333)	
Television in classroom	0.832	1.087	8.403***	0.072	1.5
	(0.781)	(0.796)	(2.398)	(0.075)	
Dropout at school	−0.221	0.332	−1.821*	0.042	0.9
	(1.019)	(1.210)	(1.023)	(0.172)	
School accesses to internet	0.289	0.090	0.642	0.075	1.6
	(0.538)	(0.572)	(1.612)	(0.157)	
Newly established school	−0.650	0.178	−0.317	0.051	1.1
	(1.472)	(1.625)	(2.291)	(0.089)	
Constant	14.159***	13.138***	12.061***		
	(2.457)	(2.710)	(3.392)		
Total				3.479	74.5
Mean predicted Vietnamese scores	20.721	20.792	16.279		
$Ln(\delta_v)$	0.468**	0.381**	0.991***		
	(0.189)	(0.174)	(0.162)		
$Ln(\delta_u)$	0.521***	0.505***	−12.735		
	(0.101)	(0.096)	(31.813)		
$Ln(\delta_\varepsilon)$	1.370***	1.371***	1.290***		
	(0.019)	(0.021)	(0.053)		
N	3218	2831	387		

Data source: 2010–2011 YLSS.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; Standard errors in parentheses; Standard errors in the decomposition results are bootstrapped with 1000 replications.

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% 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%, respectively.

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 surfer 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.

4.3.2. Decomposition results

The decomposition results show that the models can explain approximately 73% and 75% 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, which are individually responsible for around 21% and 25% of the ethnic gap, respectively. The remaining gap is attributable to parents' education, 13%, peer effect, 12%. 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%, school characteristics and peer effects, which together account for 19% 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% 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 marker 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.

5. Conclusion

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% 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 powers 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, removing the language barriers and improving school quality are more important.

Finally, given the Young Lives sample is not nationally representative, the findings here should be interpreted with caution. This limitation of the paper suggests that further studies along this line using national representative data would be useful.

Notes

1. 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).
2. The estimation results of the models with the same set of regressors are provided in Table A2 and Table A3. Most results are consistent with those from the models with the full set of regressors.

3. 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 A4 provides the estimation results of enrolment and SAGE from multilevel mixed effect Probit model. Because the estimated values of coefficients in the multi-level mixed effect Probit model are quite similar to those in Probit model (Table 4), I expect that the decomposition results derived from the two methods, if available, should be similar, too.
4. 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 (Tables 7 and 8) show that school travel time is insignificant when school characteristics are controlled for.
5. 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.
6. The decomposition for the Tobit model of SAGE shows that the total explained part accounts for 102% of the gap.
7. 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 A7 and Table A8). 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.
8. 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 A1). 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.

Acknowledgements

I thank Blane D. Lewis and Raghbendra Jha for their valuable comments. I also thank the reviewers for their constructive comments. The data used in this publication come from Young Lives, a 15-year study of the changing nature of childhood poverty in Ethiopia, India, Peru and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Department for International Development (DFID), with co-funding from Irish Aid. The views expressed here are those of the author(s). They are not necessarily those of Young Lives, the University of Oxford, DFID or other funders. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure statement

No potential conflict of interest was reported by the author.

References

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

Badiani, Reena, B. Baulch, L. Brandt, V. H. Dat, N. T. Giang, J. Gibson, J. Giles, I. Hinsdale, P. Hung, and V. Kozel. 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 Bart Cockx. 2013. "Pure Ethnic Gaps in Educational Attainment and School to Work Transitions: When Do They Arise?" *Economics of Education Review* 36: 276–294. doi:10.1016/j.econedurev.2013.07.006.

Baker, Frank B. 2001. *The Basics of Item Response Theory*. 2nd ed. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Bauer, Thomas K., and Mathias Sinning. 2010. "Blinder–Oaxaca Decomposition for Tobit Models." *Applied Economics* 42 (12): 1569–1575. doi:10.1080/00036840701721612.

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

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–846. doi:10.1086/260084.

Bollen, Kenneth A., Jennifer L. Glanville, and Guy Stecklov. 2002. "Economic Status Proxies in Studies of Fertility in Developing Countries: Does the Measure Matter?" *Population Studies* 56 (1): 81–96. doi:10.1080/00324720213796.

Cook, Michael D., and William N. Evans. 2000. "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *Journal of Labor Economics* 18 (4): 729–754. doi:10.1086/209975.

Dang, Hai-Anh. 2012. "Vietnam: A Widening Poverty Gap for Ethnic Minorities." In *Indigenous Peoples, Poverty and Development*, edited by Gillette Hall and Harry Patrinos, 304–343. Cambridge: Cambridge University Press.

Dang, Hai-Anh H., and Paul W. Glewwe. 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.

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.

Ewert, Stephanie, Bryan L. Sykes, and Becky Pettit. 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. doi:10.1177/0002716213503100.

Filmer, Deon, and Kinnon Scott. 2012. "Assessing Asset Indices." *Demography* 49 (1): 359–392. doi:10.1007/s13524-011-0077-5.

Ford, Margot. 2013. "Achievement Gaps in Australia: What Naplan Reveals About Education Inequality in Australia." *Race Ethnicity and Education* 16 (1): 80–102. doi:10.1080/13613324.2011.645570.

Glewwe, Paul, Qihui Chen, and Bhagyashree Katare. 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. doi:10.1002/app5.102.

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

GSO. 2010b. *Vietnam Household Living Standards Survey 2010*. Vietnam: General Statistics Office.

Haile, Demewoz, Dabere Nigatu, Ketema Gashaw, and Habtamu Demelash. 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. doi:10.1186/s13690-015-0113-9.

Imai, Katsushi S., Raghav Gaiha, and Woojin Kang. 2011. "Poverty, Inequality and Ethnic Minorities in Vietnam." *International Review of Applied Economics* 25 (3): 249–282. doi:10.1080/02692171.2010.483471.

Jacobson, Jerry Owen, Paul Robinson, and Ricky N. Bluthenthal. 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): 462–476. doi:10.1016/j.socscimed.2006.08.032.

Leigh, Andrew, and Xiaodong Gong. 2009. "Estimating Cognitive Gaps between Indigenous and Non-Indigenous Australians." *Education Economics* 17 (2): 239–261. doi:10.1080/09645290802069418.

Liu, Amy Y. C. 2004. "Flying Ducks? Girls' Schooling in Rural Vietnam: A Revisit." *Asian Economic Journal* 18 (3): 293–318. doi:10.1111/j.1467-8381.2004.00194.x.

McCaig, Brian, Dwayne Benjamin, and Loren Brandt. 2015. "Growth with Equity: Income Inequality in Vietnam, 2002-12." *The Journal of Economic Inequality* 15: 22–46.

Minnesota Population Center. 2017. Integrated Public Use Microdata Series, International: Version 6.5 [dataset]. Minneapolis: University of Minnesota. http://doi.org/10.18128/D020.V6.5.

Montgomery, Mark R., Michele Gragnolati, Kathleen A. Burke, and Edmundo Paredes. 2000. "Measuring Living Standards with Proxy Variables." *Demography* 37 (2): 155–174. doi:10.2307/2648118.

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

Nguyen, D., L. Parker, M. Doan, and L. Brennan. 2012. "Decision Making Dynamics within the Vietnamese Family Unit." In Proceedings of ANZMAC 2012.

Nørlund, Irene, Carolyn Gates, and Cao Dam Vu. 1995. *Vietnam in a Changing World*. Vol. 17. London: Taylor & Francis Ltd. Scandinavian Institute of Asia Studies, Studies on Asian Topics.

Paxson, Christina, and Norbert Schady. 2007. "Cognitive Development among Young Children in Ecuador the Roles of Wealth, Health, and Parenting." *Journal of Human Resources* 42 (1): 49–84. doi:10.3368/jhr.XLII.1.49.

Pham, Hung T., and Barry Reilly. 2009. "Ethnic Wage Inequality in Vietnam." *International Journal of Manpower* 30 (3): 192–219. doi:10.1108/01437720910956727.

Powers, Daniel A., Hirotoshi Yoshioka, and Myeong-Su Yun. 2011. "Mvdcmp: Multivariate Decomposition for Nonlinear Response Models." *Stata Journal* 11 (4): 556–576.

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

Ray, Ranjan, and Geoffrey Lancaster. 2005. "The Impact of Children's Work on Schooling: Multi-Country Evidence." *International Labour Review* 144 (2): 189–210. doi:10.1111/j.1564-913X.2005.tb00565.x.

Rew, W. Joshua. 2008. "Provincial, Ethnic, and Gender Disparities in Education: A Descriptive Study of Vietnam." In *Inequality in Education*, 307–323. Dordrecht: Springer.

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

Sakellariou, Chris. 2008. "Peer Effects and the Indigenous/Non-indigenous Early Test-score Gap in Peru." *Education Economics* 16 (4): 371–390. doi:10.1080/09645290802133065.

Van Der Linden, Wim J., and Ronald K. Hambleton. 1997. "Item Response Theory: Brief History, Common Models, and Extensions." In *Handbook of Modern Item Response Theory*, 1–28. New York: Springer.

Van de Walle, Dominique, and Dileni Gunewardena. 2001. "Sources of Ethnic Inequality in Viet Nam." *Journal of Development Economics* 65 (1): 177–207. doi:10.1016/S0304-3878(01)00133-X.

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

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

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