Student Paper

Young Lives

Breaking the Cycle: The Intergenerational Transmission of Human Capital

Andrew Wheeler

Thesis submitted in part fulfilment of the requirements for the degree of MSc in Economics for Development, University of Oxford.

The data used come from Young Lives, a longitudinal study of childhood poverty that is tracking the lives of 12,000 children in Ethiopia, India (in the states of Andhra Pradesh and Telangana), Peru and Vietnam over a 15-year period. www.younglives.org.uk

Young Lives is funded by UK aid from the Department for International Development (DFID) and co-funded by Irish Aid from 2014 to 2016.

Breaking the Cycle: The Intergenerational Transmission of Human Capital

Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Economics for Development at the University of Oxford

3rd June 2016

By

Andrew Wheeler

Abstract

This paper examines the causal effect of parental education on the cognitive and non-cognitive development of children. I find that a parent's education is a strong determinant of their child's verbal aptitude, numerical aptitude and educational aspirations. Parents who complete high school rather than just primary school will on average lift their children's cognitive performance by 24 percentiles in maths, 15 percentiles in vocabulary and 23 percentiles in reading tests. Children of these parents will also aspire to complete two more years of schooling. Somewhat surprisingly, I find that parental education has no impact on children's self-esteem or self-efficacy. These results are robust to various specifications. I estimate these effects using instrumental variables, taking a change in education policy with differential effects on North Vietnam and South Vietnam as my instrument. The instruments used are relevant and strong, and there is sound cause to believe that they are valid. To my knowledge, this is the first study to derive a causal relationship between parental education and non-cognitive development. It also contributes to a sparse and unsettled literature on the causal relationship between parental education and cognitive development.

I would like to acknowledge and provide my sincere thanks to the Young Lives research group for its generous support of this study, my supervisor Stefan Dercon for his valuable comments and the Rhodes Trust for its sponsorship of my studies

Table of Contents

1 Introduction

Indicators of poverty tend to be highly persistent across generations, and schooling is no exception (Pfeffer, 2008). The great challenge for development actors is to disrupt these cycles of poverty, and replace them with positive cycles of prosperity. This paper underwrites this idealistic goal with a practical policy recommendation: educate tomorrow's children by educating today's children. Better-educated parents give birth to healthier children (Güneş, 2005), give their children better at-home care (Davis-Kean, 2005), and act as positive role models (Dryler, 1988). They are a constant factor in the turbulent and unpredictable life of a child, and one that can be influenced for the better. Governments, international bodies, NGOs and other development actors should take these significant intergenerational externalities into account, and expand schooling to capture them.

Using an instrumental variable technique and data from Young lives, I find that raising an average couple's education from primary to secondary school level is expected to improve their children's cognitive performance by 24 percentiles in maths, 15 percentiles in vocabulary and 23 percentiles in reading tests. These children are also likely to be more educated themselves, as they aspire to two more years of schooling than the children of primary school-educated parents. The instruments used are valid and strong, and the results are robust to various specifications. These are significant figures, tempered by the finding that parental education does not influence children's self-esteem or self-efficacy.

The existing literature on the influence of parental education on children's cognitive and non-cognitive development is inconclusive and relatively under-developed. The few studies that attempt to derive causality disagree on whether parental education increases or has no effect on cognitive development, and to my knowledge there are no such studies on non-cognitive development. It is surprising that the literature has not focused on non-cognitive skills, such as self-efficacy and self-esteem, since they are complementary to cognitive skills and influential in their own right (Cunha et al., 2010), and are more malleable than cognitive skills throughout a person's life (Borghans et al, 2008). Furthermore, the literature is even sparser in its consideration of developing countries, which is likely to be due to data limitations. This paper is unique in

its exploration of both cognitive and non-cognitive dimensions of child development. It does so through the lens of a developing country, Vietnam, using data from the Young Lives longitudinal study.

This main contribution of this paper is to isolate the causal effect of parental education through the use of instrumental variables. Much of the childhood development literature acknowledges the potential endogeneity of parental education (Holmlund et al., 2011), particularly due to its correlation with unobserved ability. I use characteristics of Vietnamese education reforms in 1981 to design an instrumentation strategy. These reforms aimed to unify the education systems in North Vietnam and South Vietnam, which faced very different educational environments prior to the reforms. The most significant component of the reforms was an extension of the schooling system from 10 to 12 years in the North, while the 12-year system was retained in the South.

The remainder of the paper is organised as follows: Section 2 reviews the relevant literature; Section 3 outlines the identification strategy; Section 4 describes the data used; Section 5 gives the results; and Section 6 concludes.

2 Previous Literature

This paper is motivated by a growing literature on the impact of parental education on children's cognitive and non-cognitive development. Since Caldwell (1979) first identified maternal education as an important correlate of child mortality, a substantial number of studies have found a relationship between parental education and a range of indicators of early childhood development. The majority of these studies attempt to explain child health outcomes, while more recent literature takes an interest in cognitive and non-cognitive skills. This paper follows the recent literature by focusing on the latter.

Much of the recent literature is unable to explain or provide evidence into the causal links between parental education and child development. Günes (2015), Duflo (2001) and Vikram, Vannemann and Desai (2010) note the sparseness of convincing causal evidence, claiming that many studies ignore the potential endogeneity of parental education, and merely include a range of weak controls in their specifications. Recent research benefits from access to various methods and data with which to derive causality, commonly using

twin studies, adoptee studies and instrumental variables (IV). Randomised experiments are rare, given the lengthy delay between parents' acquisition of education and their children's developmental outcomes. I do not have data on twins or adoptees, so I take an IV approach.

Holmlund et al. (2011) discuss the relative designs, merits and drawbacks of twin, adoptee and IV studies in the context of the intergenerational transmission of education. Twin studies typically compare the outcomes of the children of monozygotic twins, on the assumption that such a comparison controls for the intergenerational transmission of unobservable genetic traits. This enables researchers to isolate the impact of environmental conditions on child development. Similarly, adoption studies can control for unobservable genetic effects, since adopted children's genes are independent of their parents'. IV studies take a different approach, using relevant and exogenous environmental changes to isolate the exogenous component of parental education. The literature typically uses education reforms as instruments, which has informed the choice of instrument in this study.

Aside from the usual concerns of external validity, these three methods each suffer from further econometric concerns. Twin studies do not control for heterogeneity in twins caused by early childhood outcomes, which may introduce endogeneity. For example, twin birthweights can differ substantially (Cheung et al., 1995), and low birthweight is associated with poorer development of the immune system (Chandra, 2002), which may affect both an individual's level of education and effectiveness as a parent.

With regard to adoption studies, adopted children are sometimes placed with relatives or with couples similar to their own parents, in which case estimates may be subject to selection bias. A limited number of studies control for this by studying children who have been randomly assigned to families (e.g. Sacerdote, 2007; Bjorkland et al., 2006).

Potential issues with IV studies include weak instruments, poor identification assumptions, and strong correlations between the instruments for mothers and fathers. Additionally, instruments used in this literature typically affect school choice for those with otherwise low levels of schooling, so the results should be interpreted as locally-relevant (Carneiro et al., 2013). However, this is not necessarily a major drawback, since this is often the population to which we would like to target educational reforms. Each of these methods

provides useful means through which to analyse causality. However, it is important to be mindful of these stated weaknesses when comparing studies that use different techniques, since their results often differ.

2.1 Parental Education and Child Cognitive Development

There is substantial disagreement about the impact of parental education on child cognitive development. On balance, adoption and twin studies tend to find insignificant results, although these are often based on small sample sizes, while IV studies tend to find positive and significant results. The vast majority of these studies use data on developed countries, with almost no evidence pertaining to developing countries.

Scarr and Weinberg (1978) were arguably the first to use an adoption study to assess the impact of parental education on children's IQ scores. They used data on families in the United States (US), and, like most of the subsequent literature, found no significant effect. However, they did find that the biological mother's education had a significant effect on test scores, prompting many to adopt a 'nature' over 'nurture' view of cognitive development. This paper prompted a series of similar adoption studies, many of which found a positive relationship between maternal education and cognitive development (e.g. Loehlin et al., 1989; and Capron and Duyme, 1989).

More recently, Bingley et al. (2009) used data on Danish twins to analyse the impact of maternal education on children's academic achievement. They find that maternal education has no effect, and paternal education has a marginally negative effect, on ninth grade grade-point average. They believe that the latter effect may be due to the fact that a father's working hours are likely to increase with education. Hægeland et al. (2010) return similar results for twins using a Norwegian census level dataset; they find that a mother's education does not impact her child's schooling outcomes. However, using data on adopted children, they find that maternal education has a positive impact on children's school performance. Paternal education is found to be insignificant in both cases. They provide no convincing explanation for these results, as better educated mothers' longer working hours have no detrimental effect on children; in fact, they tend to spend more time with them.

In contrast to these studies, Sacerdote (2000) finds that both maternal and paternal education have a small positive effect on child cognitive development, using data from Britain and the United States on adopted

children. Similarly, using data on adopted children in the US, Neiss and Rowe (2000) find that average parental education has a modest positive impact on children's verbal intelligence.

IV studies typically find stronger evidence that parental education increases child cognitive development. For instance, Carneiro et al. (2013) instrument maternal education with US schooling costs, local labour market characteristics and distance to college during the mother's adolescence, and find a positive relationship between maternal education and maths and reading test scores, and a negative relationship with grade repetition. Maurin and McNally (2008) use the 1968 student riots in France, which dramatically increased tertiary education pass rates, as an instrument for paternal education. They show that increased paternal education results in improved school performance, measured by grade repetition. Oreopolous, Page and Stevens (2006) also assess parental education's effect on grade repetition, using a change in compulsory schooling laws in the UK. They find that an increase in either parent's education results in a reduction in the likelihood that their child repeats a grade.

2.2 Parental Education and Child Non-Cognitive Development

The literature on the relationship between parental education and non-cognitive development is far weaker than that which looks at cognitive development (Heckman and Rubinstein, 2001). In fact, I am not aware of a study that claims to establish a causal link between the two factors. This is surprising, given non-cognitive skills have been studied for decades (e.g. Bowles and Gintis, 1967; Jencks et al., 1979), and environmental characteristics have been found to influence them (Heckman, 2007). Additionally, non-cognitive skills are known to increase the returns to investment in a wide variety of educational programs (Cunha et al., 2010), and are more malleable than cognitive skills throughout a person's life (Borghans et al, 2008). Several papers have shown that early nurturing of non-cognitive skills is a cost-effective way to improve future outcomes (e.g. Bowles, Gintis and Osborne, 2001). As a result, it is clear that this paper addresses an important and under-researched question.

While the literature does not offer much guidance, on balance it seems to suggest a positive relationship between parental education and children's non-cognitive development. Possibly the most relevant paper to this question is by Dercon and Krisnan (2009), who find a positive correlation between a caregiver's education and a range of non-cognitive development measures. However, they do not make any claims of causality. The

general notion of parenting is also linked to non-cognitive development in numerous studies (e.g. Sylva et al, 2004; Masten and Shaffer, 2006; Ermisch, 2008). However, the literature is not conclusive; for example, Coneus, Laucht and Reuß (2012) find that parental investments do not influence children's emotional skills.

2.3 Gaps Addressed by this Paper

It is clear that the literature is far from conclusive: the evidence on cognitive development is conflicting; studies on developing countries are sparse; and there is no evidence on the causal impact of parental education on non-cognitive development. Moreover, no existing studies focus on causal effects in Vietnam, which could yield interesting results given its female gender bias in education (Dercon and Singh, 2013). This paper addresses these gaps by using Vietnamese education reforms implemented in 1981 as an instrument for parental education, and matching this with detailed data from the Young Lives longitudinal survey.

3 Identification Strategy

3.1 Conceptual Framework

Most studies on the determinants of child development base their analysis on a framework of proximate causes. Strauss and Thomas (2007) provide a good summary of these models, which are typically based on a health production function such as:

$$H = H(N; A, B_H, D, \mu)^1$$

However, such a model cannot be used to assess the impact of parental education on child development, because parental education can influence several proximate causes of development. For example, if we control for health inputs in our regression then we are likely to understate the impact of parental education on child development, since a parent's education may influence the volume and nature of health inputs given to their child.

[1] H represents health outcomes, N represents health inputs and behaviours, A represents socio-demographic characteristics, B_H represents relevant dimensions of family background, D represents relevant environmental characteristics, and μ represents errors

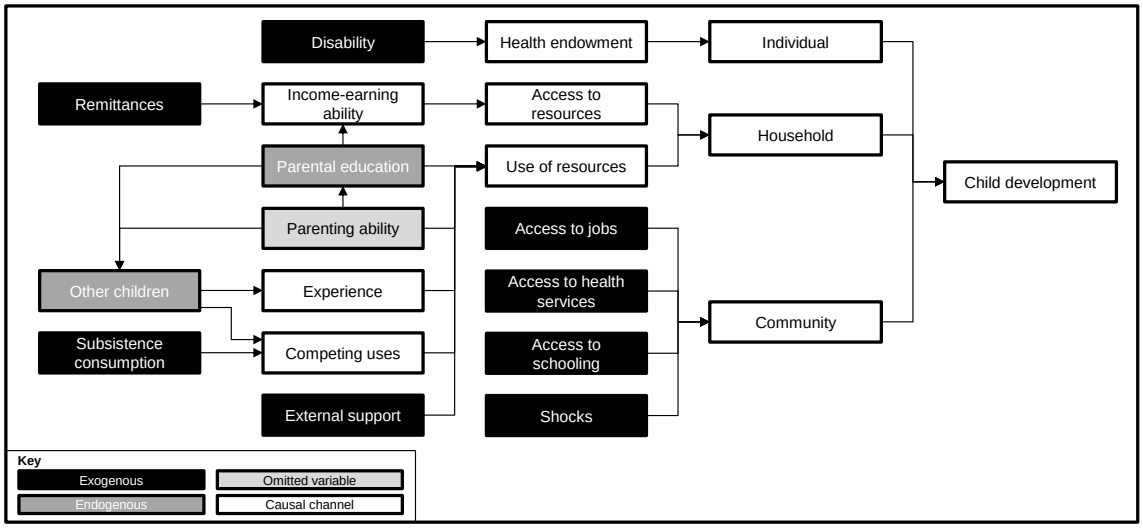
We therefore need an alternative framework that more comprehensively captures the impact of parental education. I propose a framework that is based on the observation that, holding the country constant, a child's health will differ on the basis of three distinct dimensions: individual, household and community. This framework allows us to separate the effects of parental education from the proximate causes of child health, such as nutrition, and leads to a reduced form equation of the form:

$$H_i = H_i(\boldsymbol{I}_i, E_i, \boldsymbol{HH}_i, \boldsymbol{C}_i)$$ [2]

where $\boldsymbol{I_i}$ represents a vector of individual characteristics; $\boldsymbol{E_i}$ represents parental education; $\boldsymbol{HH_i}$ represents a vector of exogenous household characteristics that are not affected by education; and $\boldsymbol{C_i}$ represents a vector of community characteristics.

Figure 1 outlines one possible characterisation of this framework.

FIGURE 1: DRIVERS OF CHILD DEVELOPMENT



I rely on this framework, and the assumption of linearity for simplicity, in the following model used in this paper's empirical analysis:

[2] Bold text indicates a vector

$$H_i = \beta_0 + \beta_1 E_i + \boldsymbol{\beta_2 I}_i{}^3 + \boldsymbol{\beta_3 HH}_i{}^4 + \boldsymbol{\beta_4 C}_i{}^5 + \varepsilon_i$$

This allows me to estimate the effect of parental education, controlling for variables that may affect child development but that are unlikely to be channels through which parental education affects child development.

This specification may plausibly suffer from heteroskedasticity. For example, if education is necessary but not sufficient for child cognitive development, then poorly-educated parents will tend to have similarly poorly-performing children, while we will see more variance among the children of otherwise-heterogeneous well-educated parents. Consequently, I use heteroskedasticity-robust standard errors.

3.2 IV Strategy

3.2.1 Overview

A simple regression of parental education on early childhood development is likely to suffer from endogeneity, since parental education may be correlated with a host of unobservable characteristics that influence child development. For example, parental education may be correlated with unobservable parenting ability, which influences child cognitive development. As a result, I use an instrument that captures exogenous variation in parental education due to educational reforms implemented in Vietnam in 1981. The primary effect of the reforms was to introduce a 12-year secondary education system in the North.

3.2.2 Vietnamese Educational Reforms

In 1979, a new Vietnamese education policy was drafted and by the 1981-82 school year it had begun to be implemented. These reforms were ambitious in scope, and were designed to align the education systems of the North and the South. Among the various reforms, the movement towards a common 12-year curriculum was perhaps the most significant. This extension of the schooling system to 12 years affected the North and the South differentially; before the reforms, the South already had a 12-year education system, whereas the

[3] I use the existence of a long term health problem as a control

[4] I use the following variables as controls: whether a child has a long-term health problem; whether the family receives remittances; whether any non-parents within the household provide financial support to the child; and whether both parents live in the household

[5] I use the following variables as controls: average daily wage of a municipal unskilled worker; number of nurses in the community; and the average daily wage of a school teacher relative to a municipal unskilled worker. The final control is used as a proxy for access to schooling, since there is no variation in access to primary schools and very little variation in access to secondary schools. I expect that teacher's relative wages are an indication of the scarcity of teachers, and therefore represent difficulty of access to schooling

North had a 10-year system. The reforms were implemented gradually, with the new curriculum introduced to the 1981 cohort in each successive year. The 12-year duration of schooling was also phased in; for instance, in 1988 the education system in the North was still only 11 years (Marr and White, 1988). Consequently, the cohort of students entering school from 1981 onwards was fully exposed to the reforms, while those students already in school in 1981 were only partially exposed. These are the sources of exogenous variation that I exploit in my instrument.

I assume that, in the absence of the reforms, the North and the South would have followed the same trajectories, sustaining the difference in schooling outcomes between the two regions. The results in Section 0 support this assumption.

3.2.3 Instrument Design

The reforms influence the difference in education between the North and the South more strongly than the level of education in both regions. As a result, I restrict my sample to individuals in the North and use education in the South as a baseline upon which to add the effect of the reforms. Therefore, the education of an individual born in a given year can be estimated using the average education for a similar person in the South and the individual's exposure to the reforms. An individual's exposure to the reforms is jointly determined by whether she entered school from 1981 onwards and, if she was enrolled before 1981 and was still enrolled in 1981, the number of years of non-tertiary schooling she still had the opportunity to complete.

I therefore use four instruments to jointly explain aggregate parental education. The first is the average across both parents of: the average education level in South Vietnam for mothers / fathers of the same age, calculated separately for those with and without other children. The others are: an indicator for whether both parents were fully exposed to the reforms; an indicator for whether just one parent was fully exposed to the reforms; and, if either parent was partially exposed to the reforms, the average number of years of non-tertiary schooling the parents still had the opportunity to complete upon the introduction of the reforms. This can be represented as:

$$E_i = \alpha_1 + \alpha_2 AvEdSV_i + \alpha_3 BothFull_i + \alpha_4 OneFull_i + \alpha_5 RemEd_i + \nu_i$$

The reforms are likely to have had the greatest impact on students who would have completed at least 10 years of education under the former system, so it could be argued that the results should be interpreted as

a local average treatment effect for these individuals. However, it is reasonable to believe that there are a number of other channels through which the effects could be felt by other groups. For example, students may drop out when becoming 'close enough' to the final year, social expectations of a person's education level may rise, and, since others are more educated, the returns to lower levels of education may fall. These possible channels are also likely to more strongly affect those with higher levels of education, so I interpret the results to be most relevant to children of better-educated parents.

It is important to acknowledge that the unbiasedness of 2SLS is an asymptotic result, and that multiple instruments may exacerbate finite sample size bias (Bound, Jaeger and Baker, 1995). While this may be an issue, the sample used in this instance is relatively large, and including additional instruments improves the asymptotic efficiency of 2SLS methods, so I proceed with this approach.

3.2.4 Instrument Validity

There are two key concerns that I need to satisfy for these instruments to be valid. The first is that they do not directly affect child development in any way other than through parental education, and the second is that the impact of the reforms is not correlated with other determinants of child development. The first is easily satisfied; the reforms only targeted schooling, and were not directly associated with health, employment or other relevant reforms. It is also unlikely that the education system in the South has a causal effect on the education levels of those in the North in the same cohort.

The second concern is less easily satisfied. Since an individual's exposure to the reforms is determined by her age, one might be concerned that any instrument will merely reflect the effect of age on education. There may have also been changes in regulations, community characteristics or household decisions that were unrelated to the reforms, but are nonetheless correlated with them. I am primarily concerned with migration patterns, health reforms and family size.

This is a cohort study, so younger mothers are likely to have less education, meaning that the instrument may merely reflect the study design and not any causal pathway. Further, an individual's age may have independent effects on child development, violating the validity condition. This study is designed to account for these concerns. Firstly, the instrument explains the difference in education between North and South Vietnam, which is unlikely to be related to age except through the education reforms. Secondly, age is not

expected to have material independent effects on child development. To test this, I look at arguably the most plausible channel through which age could independently affect test scores: premature birth. I find no statistical relationship between mother's age and premature birth, as per Table 11. This conclusion is supported by Reichman and Pagnini (1997), who show that adverse birth outcomes are not determined by age, but rather by the socio-economic characteristics of younger mothers.

Migration patterns will be of concern if their effect on a migrant's child's development follows a similar pattern to the education reforms' impact on parental education. For this to be the case it is necessary, but not sufficient, for there to be a discontinuity in migration patterns for children whose parents entered school from 1981 onwards. Anh, Tacoli and Tanh (2003) describe migration patterns in Vietnam following unification, and do not provide any evidence to support this possibility. They explain that between 1975 and 1986, population movements were tightly controlled and focused on urban-rural and rural-rural migration, and that in 1986 *Doi Moi* was introduced, which liberalised migration flows. There is little reason to believe that there would be a material discontinuity in the effects of these policies on the basis of an individual's age.

If health reforms affect the validity of the instrument, we would expect to see a shock to the health outcomes of children entering school from 1981 (i.e. born after 1974) that affected the North and South differentially. One possible source of such a pattern could be the end of the Vietnam War in 1975. However, Banister (1985) shows that post-war changes in life expectancy and child mortality were relatively uniform across the North and the South, supporting the instrument's validity.

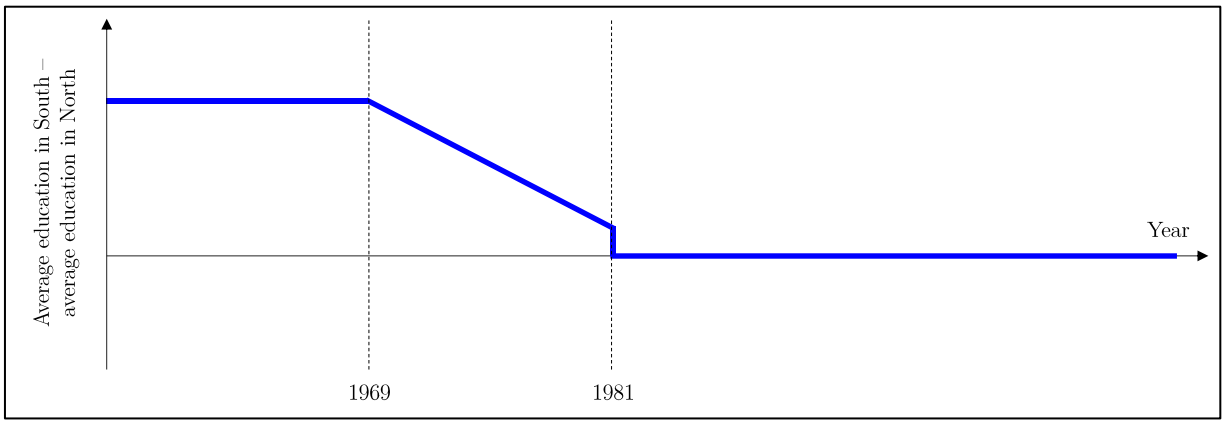
Finally, the instrument may affect family size because being born after 1981 indicates that a mother is younger and is likely to have fewer children. Family size may also be affected by the same unobservable variables that cause endogeneity in parental education, such as parenting ability. Following Acemoglu and Angrist (2000), I account for this threat to instrument validity by treating family size as endogenous. I use the same instruments as I do for parental education, and find that they are relevant and that the rank condition holds. Following Angrist and Pischke (2010), I do not consider endogeneity in other variables that pose no threat to the validity of my instruments.

3.2.5 Instrument Relevance

The instruments used must also explain some of the variation in education levels after controlling for other exogenous variables. Table 10 shows the results for the first stage regressions of the instruments on average parental education. As expected, each of the instruments is a statistically significant correlate of average parental education, and the null of irrelevance in underidentification tests is rejected in every 2SLS regression performed, confirming that the instruments are relevant.

To test whether this effect is driven by reforms, I use two 'artificial reforms' five years either side of 1981. I construct instruments for these reforms in the same way as for the actual 1981 reforms. Given the expected impact of the reforms is a smooth narrowing of the education gap between North Vietnam and South Vietnam, creating an 'artificial reform' in a year other than 1981 will weaken the instruments, but not necessarily render them completely irrelevant.

FIGURE 2: EXPECTED EDUCATION GAP BETWEEN NORTH VIETNAM AND SOUTH VIETNAM



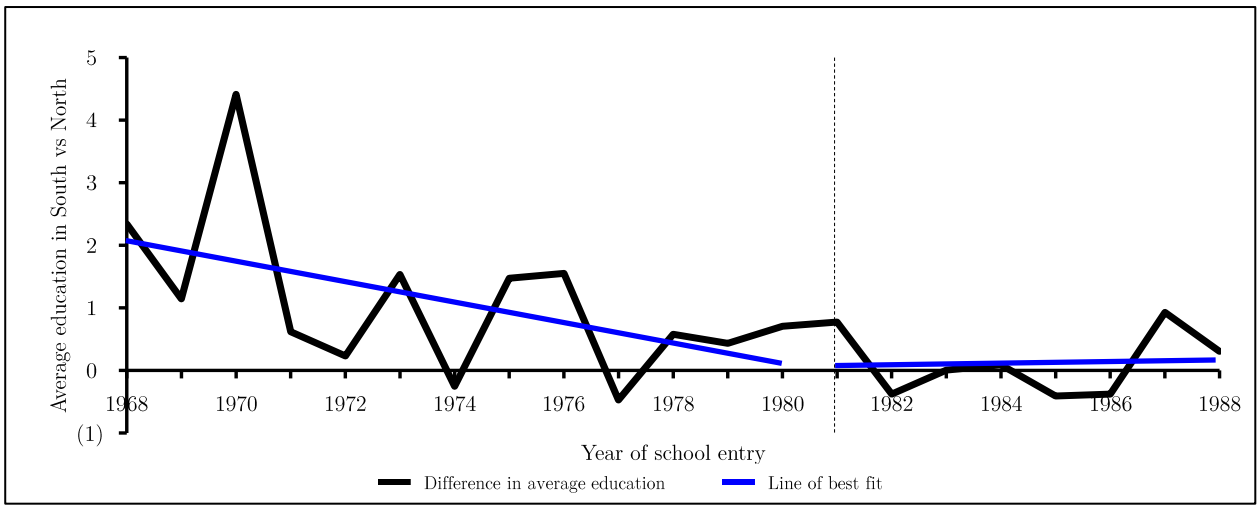
From Figure 2, we would expect that an placebo reform year after 1981 will pick up very little variation in the education gap, while a placebo reform year before 1981 will pick up some of the variation due to partial exposure to the reforms, so may return significant results, but they will be less significant than those when 1981 is used as the reform year. In the case of the siblings, the instruments pick up both the impact of the reforms and the impact of the age of the parents. As a result, placebo instruments are likely to remain significant. This is what we observe, as shown in Table 14. These results confirm that the reforms are likely to be the driver of the change in the education gap.

3.2.6 Testing Common Trends Assumption

The IV methodology relies on the assumption that the difference between education in North and South Vietnam would have remained constant. Figure 3 shows the observed difference, which takes the impact of the reforms into account. Unfortunately data are not available for enough pre-reform years (pre-1969), but the figure suggests that the difference in the post-reform years is relatively constant. I test this hypothesis by regressing the difference in education against time in two groups: pre-1981 and post-1981.

Table 13 shows that year of school entry is not statistically significant post-1981, supporting the hypothesis that education trends were similar in the North and the South. While this does not necessarily indicate that the trends would have been similar in the absence of the reforms, it is reasonable to believe that this is the case, since the factors that would have caused the two systems to diverge in the absence of the reforms are likely to have had a similar effect in the presence of the reforms.

FIGURE 3: DIFFERENCE BETWEEN AVG EDUCATION IN NORTH AND SOUTH VIETNAM BY YEAR OF SCHOOL ENTRY



4 Data Characteristics

4.1 The Sample

The data used in this study are from the Young Lives Programme, funded by DFID and based at the University of Oxford. Young Lives is a longitudinal study based in four countries, including Vietnam. In each

country, data are collected for two cohorts: 2,000 children in a "Younger Cohort" (aged 6-18 months in 2002) and 1,000 children in an "Older Cohort" (aged 7.5-8.5 years in 2002). The study now spans 4 rounds and 15 years.

In Vietnam, Young Lives gathered data from 20 sites, with four sites selected in five provinces. The study purposefully over-sampled poorer communities, with 48 percent of sites classified as poor, 29 percent as average and 23 percent as above-average (Young Lives, 2011). Within each site, children were randomly selected. The households surveyed are therefore not nationally representative: they are less likely to own property, own fewer other assets, and more likely to be classified as poor by their local authorities (Nguyen, 2008).

The study consists of a child questionnaire, a household questionnaire and a community questionnaire. Across these three questionnaires, data were collected on topics such as: child health, cognitive achievement, attitudes and aspirations; household composition and assets; access to basic services; and economic activity in the community. The longitudinal nature of the study allows us to observe how these characteristics change over time.

Attrition rates for the study are very low. In the Younger Cohort, this study's cohort of interest, attrition is only 3.6% (Young Lives, 2013). The cause of attrition tended to be migration (2.2%), with small numbers due to refusal (0.6%). In addition to attrition, 0.6% of the cohort died between rounds 1 and 4.

Table 3 presents a range of summary statistics for the Younger Cohort.

4.2 Measurement of Variables

Young Lives collects data that can be used to measure cognitive and non-cognitive development in rounds 3 and 4. Non-cognitive development is assessed by asking children how strongly they agree or disagree with a series of statements, and is measured on a 4-point Likert scale. These statements were based on those found in the educational psychology literature, and were adapted for use with Vietnamese children. I construct indicators for self-efficacy, self-esteem and aspirations with these responses using a similar method to that used by Dercon and Sanchez (2013). Self-efficacy refers to a child's sense of control over his / her life, and I

calculate its indicator using responses to five questions[6]. The indicator for self-esteem was calculated using five questions in round 3[7] and six questions in round 4[8] that measured a child's level of pride or shame in various dimensions of their life. Aspirations in round 4 are measured by the number of years of education a child would like to complete.

Cognitive development is measured based on a child's responses to three aptitude tests. Verbal ability was measured using the Peabody Picture of Vocabulary Test and the Early Grade Reading Assessment, while mathematical ability was measured using a mathematics test.

4.3 Data Issues Influencing Inference

While the data and econometric approach used in this paper are generally well-suited to the question being addressed, it is important to consider potential issues that could affect inference. The issue most commonly found in the literature is that maternal and paternal education are often highly correlated. This phenomenon is known as assortative matching. This issue is often dealt with by taking the aggregate of both parents' education, as in Oreopoulos, Page and Stevens (2006). However, this approach cannot separately identify the impact of maternal and paternal education. Brierova and Duflo (2004) solve this issue by using a separate instrument for the difference between maternal and paternal education. In the absence of such an instrument, I take the average parental education level as the explanatory variable of interest.

The remaining data issues that are considered in this paper are not critical for estimation, so I use them to check the robustness of the simplified methodology. A key issue is that Young Lives does not have data on the community in which a person obtained their education. In the simplified model I assume that the region in which a person lives (North or South) corresponds with the region in which they received their education. However, given the significant amount of internal migration in Vietnam, it is likely that this assumption does not hold. If this migration is random, or at least unrelated to parental education, then results will be unbiased

[6] The questions asked children to indicate their level of agreement with the following statements (or disagreement when responses are framed in the negative): "Other people in my family make all the decisions about how I spend my time"; "I have no choice about the work I do - I must do this sort of work"; "If I try hard, I can improve my situation in life"; "I like to make plans for my future studies and work"; and "If I study hard at school I will be rewarded by a better job in future"

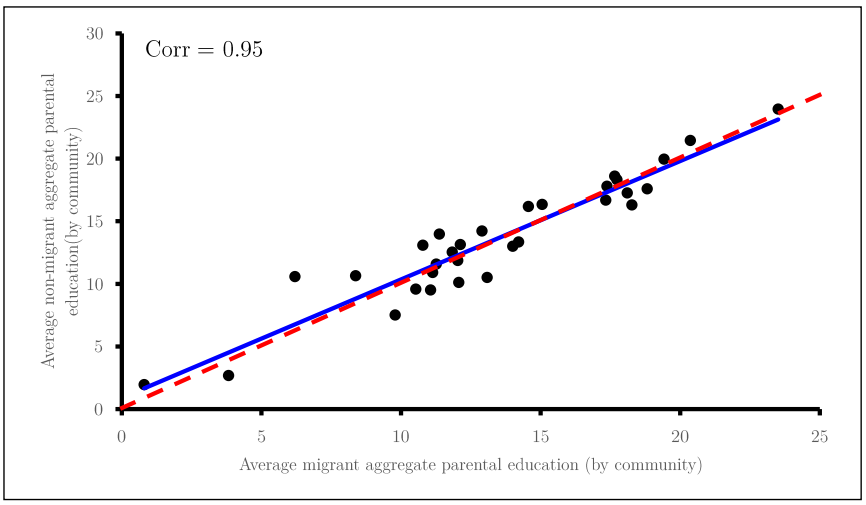
[7] The questions asked children to indicate their level of agreement with the following statements: "I'm proud of my clothes"; "I am proud of the work I have to do"; "I am never embarrassed because I do not have the right books, pencils and other equipment for school"; "I am proud of my shoes or of having shoes"; and "I am proud that I have the correct uniform"

[8] The indicators were the child's level of agreement with the above statements, as well as "Overall, I have a lot to be proud of"

and consistent, but with larger standard errors. This is likely to have been the case until the introduction of Doi Moi in 1986, as migration was largely controlled by the government (Anh, Tacoli and Thanh, 2003). However, by the mid-1990s spontaneous migration had become more common than organised migration (Hardy, 2000), which presents the possibility that migration is related to economic and social opportunities that affect child development.

I find that migrants are very similar to non-migrants, and conclude that migration patterns in the communities analysed are unlikely to be a threat to identification. Figure 4 shows that there is a very significant and near-uniform relationship between the average education of migrants and non-migrants. While we may see these results because migration leads to sorting on the basis of characteristics such as education, this may also be because identification is not significantly affected in the event of net out-migration, since such regions may be less attractive, and thus less likely to attract economic migrants from far away. Lao Cai and Hung Yen, the two Northern provinces included in Young Lives, both experience net out-migration (Anh, Tacoli and Thanh, 2003). Additionally, it is not an issue if migration occurs within the North and within the South, but not between the two. It is likely that this is largely the case. While I do not see migration as posing a threat to identification, I test the robustness of my findings to this assumption through two checks: including migrant status as a control; and restricting my sample to individuals who have resided in their community since before Doi Moi in 1986. However, the latter is likely to be highly conservative and substantially affect inference, since it reduces the sample size by approximately half.

FIGURE 4: RELATIONSHIP BETWEEN MIGRANT AND NON-MIGRANT EDUCATION



There are also a number of instances in which it is not possible to observe data of interest. These include that: data are missing for parents who do not live with the child; years of schooling is not observable for individuals who have completed certain qualifications, such as adult literacy training; years of education is only available in the second round; and the study experiences attrition.

Missing data for absent parents is a potential concern because it is likely to be due to parental death, divorce, or economic migration. These factors are almost certainly correlated with parental education and measures of early childhood development, so will introduce bias. In my simplified model I control for this by including a dummy variable for whether both parents live in the household. To test the robustness of these results, I restrict the sample to children who live with two parents, which represents a locally-relevant effect. This is a conservative approach, since only 17 fathers and 5 mothers are missing from the sample in round 1.

I do not treat the exclusion of individuals with adult education or religious education as a problem for identification, since there are only seven parents for whom this applies. Similarly, I do not account for issues associated with attrition, since it is only 3.6% across the four rounds, or for the use of education data from the second round, since only 45 parents are still in full time education in the first round.

There are some instances in which variables may be measured with error, such as parental education. While an individual's own education level is likely to be known with relative certainty, others' education may be reported with error, which will result in attenuation bias. It is possible to use reported education in another round as an instrument for parental education, however this will merely replace attenuation bias with selection bias, since education is not available for a very substantial number of parents in round 3 or 4, and the parents for whom education data are missing tend to be less educated. As a result, I do not correct for attenuation bias and acknowledge that OLS results may be biased downward.

While the issues detailed above are inconvenient, I do not see them as major obstacles to conducting inference. However, if I were to adapt the Young Lives survey to better suit this study, I would choose to include more information on caregivers, and survey multiple household members for the same information. I would prioritise additional information on each household member's town or origin, and double-reporting of the education of each household member. This would allow me to more accurately connect the reforms to parents, and deal with possible measurement error.

5 Estimation

5.1 Preferred Estimator

In this section, I outline a framework for choosing the preferred estimator for each variable of interest, and discuss which estimator I have chosen to use.

The framework relies on the following characteristics of OLS and 2SLS estimators. If parental education is exogenous, then OLS should be used as it is more efficient than 2SLS. However, if parental education is endogenous, then 2SLS is the preferred estimator if the instrumentation strategy is strong and valid. Controls should be included if one suspects conditional validity, but excluded if controls are channels through which parental education affects child development. Finally, if instruments are weak or invalid, then 2SLS may contain significant bias and neither OLS nor 2SLS will yield useful results. Although weak instruments may not be a significant issue if 2SLS results are similar to those of median-unbiased estimators. This is summarised in Table 1 below.

TABLE 1: PREFERRED ESTIMATOR DECISION FRAMEWORK

	Exogeneity	Validity	Conditional validity	Weak instruments	Preferred estimator
(1)	✓	----------	N/A	----------	OLS
(2)	×	×	×	N/A	Neither OLS nor 2SLS is suitable
(4)	×	×	✓	×	2SLS with controls
(5)	×	×	✓	✓	2SLS with controls if results are similar to median-unbiased estimators. Otherwise neither OLS nor 2SLS is suitable
(6)	×	✓	N/A	×	2SLS with no controls
(7)	×	✓	N/A	✓	2SLS with no controls if results are similar to median-unbiased estimators. Otherwise neither OLS nor 2SLS is suitable

While it is not possible to conclusively claim that any of these scenarios holds, I take (4) as the most likely scenario, and use 2SLS with controls as my estimator of choice. I present full results of OLS and 2SLS with and without controls in the Tables section of the Appendix.

First of all, I consider exogeneity. Hausman (1978) provides a useful test of this which can help guide the discussion of exogeneity, which is roughly based on a comparison of OLS and 2SLS estimates. However, it is by no means conclusive or a substitute for economic reasoning. In all cases except the Round 3 Vocabulary score, the test fails to reject the null of exogeneity. There are many reasons why this might be the case, even

if parental education is endogenous. One possible reason is that the 2SLS estimates are biased in the same direction as OLS, and another is that variance in the estimates may disguise economically-meaningful differences in the coefficients. I choose to trust in the economic logic underpinning the assumption of endogeneity, and assume that parental education is endogenous to each dependent variable.

Validity is discussed in detail in Section 3.2.4, and there are sound reasons to believe that it holds. However, there are some patterns in the results that could be explained by conditional or unconditional invalidity. For example, in many cases overidentification tests fail in the absence of household or community controls. However, this does not necessarily indicate invalidity; the five instruments have been designed to capture different dimensions of the education reforms, and some are weaker predictors of parental education than others. Therefore the difference between 2SLS estimates based on different subsets of instruments may be driven by weak individual instruments and finite sample bias, rather than invalidity.

Another potential indicator of invalidity is that estimates change when controls are added. This may suggest that there is invalidity, or it may be that the controls are channels through which parental education affects child development. While efforts have been made to choose controls that are not determined by parental education, it is likely that parental education has some influence on them. However, the estimates change in an unexpected direction when controls are added; one would expect the influence of parental education on child development to decrease when positive channels of influence are controlled for, but in these results they increase. On possible explanation for this is that the number of siblings is endogenous and a threat to validity, as described in Section 3.2.4. Therefore I take a conservative view that the instrument may not be valid without controls, but is valid with controls.

Finally, I find that instruments are strong with individual, household and community controls, but are weak when I include region fixed effects. Again, this may reflect the correlation between parental education and region; for example, better-educated parents may be able to afford housing in regions with more schools and hospitals. Alternatively, given region fixed effects are relatively demanding on the data, weak instruments with their inclusion may indicate issues with power rather than validity. However, 2SLS estimates are very similar to LIML estimates, which are median-unbiased, so weak instruments may not be an issue in this case (Angrist and Pischke, 2010).

For these reasons, I take 2SLS with individual, household and community controls as the preferred estimator, and test whether results differ when controls are added or OLS is used.

5.2 Summary of Results

5.2.1 Method for Testing Significance

This paper tests 11 hypotheses, each of which relates to the impact of parental education on a measure of cognitive or non-cognitive development. If we assume that these hypotheses are independent and that none are truly significant, then, at a 10% significance level, we would expect to see one false positive in these results. As a result, I test these hypotheses against the false discovery rate (FDR), following Benjamini et al. (2006). I use an FDR of 0.05 so that the expected number of false discoveries is less than one.

To keep the FDR less than $q = 0.05$, I sort the p-values $p_1 \leq \cdots \leq p_{11}$, find $k = \left\{\max i \,\middle|\, p_i \leq \min\left(\frac{iq}{\widehat{m_0}}, q\right)\right\}$ and reject hypotheses $p_1 \leq \cdots \leq p_k$, if any such hypotheses exist. $\widehat{m_0}$ is the estimated number of true hypotheses, found by comparing the p-values against a significance level of 0.05.

As per Section 5.1, my preferred specification is 2SLS with individual, household and community controls. The summarised results are given in Table 2 below alongside the corresponding results for OLS.

TABLE 2: SUMMARY RESULTS

Variable	**Mean**	**2SLS**					**OLS**				
		Coefficient	p-value	$iq/\widehat{m}_0$	Reject null?	Robust*	Coefficient	p-value	$iq/\widehat{m}_0$	Reject null?	Robust*
Cognitive											
R3 Maths	297	3.16	0.000	0.036	✓	✓	2.56	0.000	0.045	✓	✓
R4 Maths	46.7	3.43	0.000	0.036	✓	✓	2.59	0.000	0.045	✓	✓
R3 Vocab	303	3.15	0.000	0.036	✓	✓	2.14	0.000	0.045	✓	✓
R4 Vocab	75.1	2.13	0.000	0.036	✓	✓	1.60	0.000	0.045	✓	✓
R3 Reading	297	2.06	0.003	0.050	✓	✓	2.26	0.000	0.045	✓	✓
R4 Reading	46.3	3.24	0.000	0.036	✓	✓	2.46	0.000	0.045	✓	✓
Non-cognitive											
R3 Self-Efficacy	3.65	0.005	0.823	0.050	×	✓	0.020	0.000	0.045	✓	✓
R4 Self-Efficacy	3.64	0.023	0.256	0.050	×	✓	0.022	0.000	0.045	✓	✓
R3 Self-Esteem	3.81	0.013	0.572	0.050	×	✓	0.013	0.020	0.050	✓	✓
R4 Self-Esteem	3.39	0.013	0.459	0.050	×	✓	0.008	0.069	0.050	✓	✓
R4 Aspirations	13.2	0.26	0.003	0.050	✓	✓	0.19	0.000	0.045	✓	✓

* Indicates whether coefficients remain significant or insignificant after robustness checks

The instrumentation strategy used to obtain these findings is strong and highly relevant. As a result, there is unlikely to be significant bias in the 2SLS results.

5.2.2 Cognitive Development

There is strong evidence to suggest that parental education improves children's cognitive development: parental education has both a significant and material effect on each measure of cognitive development. Using results from round 4 for ease of interpretation, the difference between parents completing the minimum number of years of schooling (5 years) and finishing secondary school (12 years) corresponds with an improvement in children's performance of 24 percentiles in maths, 15 percentiles in vocabulary and 23 percentiles in reading. The significance of these results is robust to the inclusion of controls and further robustness checks detailed in Section 4.3.

The 2SLS results tend to be higher than OLS, which is a common result in IV studies (Li and Luo, 2004). There are several possible reasons for this, including that: measurement error results in attenuation bias in the OLS estimates; IV results are applicable to a subset of the sample, and there are heterogeneous returns to parental education; and the instrument is invalid.

Given one member of the household reported on the level of education of all other household members, it is likely that education is measured with error. This will bias OLS results towards zero. If the instruments are uncorrelated with this error, then 2SLS estimates will be higher than OLS estimates if attenuation bias is larger than endogeneity bias. This is a plausible explanation for the observed difference in coefficient magnitudes.

2SLS may also represent a local average treatment effect, and returns for the relevant subset of individuals may be higher than for the rest of the population. In this case, the local population is comprised of parents with higher levels of education. While it is likely that returns to education are lower in this population across many dimensions, the returns for children's cognitive development may be higher. Children may not benefit from incremental parental education when parents still have rudimentary knowledge of mathematics or language, but a more advanced understanding of these topics may equip parents with the skills to become better at-home teachers.

Finally, it is possible that the instrument is invalid, and that weak instruments exacerbate endogeneity bias. Unfortunately there is no conclusive test for validity, so I rely on the logic detailed in Section 3.2.4 to justify IV strategy.

Additionally, estimates vary based on which set of controls are used. Again, this may suggest that instruments are invalid, but it may also indicate that the instruments are correlated with controls and are conditionally valid. For example, communities with higher incomes and greater provision of education are likely to contain better-off parents who give birth at an older age. These parents may therefore be less exposed to the reforms than younger parents. Controlling for community characteristics accounts for this.

5.2.3 Non-Cognitive Development

It is surprising to find that parental education has a small and statistically insignificant effect on the self-efficacy and self-esteem of children; taking the 2SLS estimator, the estimates are indistinguishable from zero. Even if we believe the OLS estimates, which are very similar in magnitude to the 2SLS estimates, the approximate impact of 10 years of parental schooling is to turn a single response across five questions from 'agree' to 'strongly agree'. While the OLS estimates are statistically significant, they are economically insignificant. Possible explanations for these results include that better-educated parents have higher expectations for their children, better-educated parents exert more control over their children, or children exhibit adaptive preferences. Alternatively, the results may reflect the fact that 2SLS is less efficient than OLS.

Unsurprisingly, I find that higher parental education increases children's educational aspirations. When parents complete secondary school, on average their children desire two more years of education than the children of parents who have only completed primary school. This is remarkable when we consider that children in the sample have fairly uniformly high educational aspirations. However, it is worth noting that this is only one measure of aspirations, and it is more closely related to parental education than other measures of aspirations, so one must take care when considering these results in the context of broader aspirations.

5.3 Robustness Checks

From Section 4.3, the two key sources of potential bias come from migration and missing data on parents, and I re-run my results in three ways to test robustness to these factors. Firstly, I restrict my sample to children with both parents; secondly, I include migrant status as a control; and thirdly, I restrict the sample to non-migrants.

I find that the results in these robustness checks are largely consistent with those in my main specification, and conclude that my results are robust to these possible sources of bias. However, restricting the sample to non-migrants leads to some sources of inconsistency, as the 2SLS cognitive results become insignificant. This may be because the sample is much smaller, reducing power and exacerbating problems with weak instruments. As mentioned in Section 4.3, this was a highly conservative robustness check, since we are only concerned with migration between the North and the South, and the data measure total migration, including between neighbouring villages.

An additional robustness concern stems from the fact that the instruments become weak in the presence of region fixed effects. I consider this to primarily be a symptom of low power due to the statistically-demanding nature of region fixed effects. When community controls are included that take into account local incomes, health services and educational services, the instruments are relatively strong and the first stage regression is highly relevant. As a result, I do not consider weak instruments in the presence of region fixed effects to be of major concern.

6 Discussion

Childhood cognitive and non-cognitive development is known to be an important driver of adult development and wellbeing. This paper finds that parental education can play an important role in shaping several dimensions of these measures. Using instruments to capture the effect of exogenous education reforms in Vietnam, I find a robust, positive causal relationship between average parental education and mathematical aptitude, verbal aptitude and educational aspirations. These results have important policy implications, as they highlight an intergenerational channel through which education can tackle poverty, and provide evidence

that cycles of poverty can be broken. Moreover, it points to a potential under-valuation of education by national governments that only consider the returns to education for the current generation.

The results for measures of self-efficacy and self-esteem are less conclusive, but point to an economically-insignificant effect. While it is possible that there is not enough power to claim statistical significance, the effect size is small and unlikely to shape educational policy. More research is required to support or refute these claims.

This paper presents numerous avenues for future research. Given the sparseness of existing literature, more research is required into the questions addressed by this paper, particularly in developing country contexts. This is likely to require additional sources of data and innovative identification techniques, given endogeneity concerns and the fact that this paper identifies local effects. Additionally, it is important to understand the channels through which parental education affects child cognitive and non-cognitive development. The policy implications are very different if, for example, it is through increased income, compared with an improved ability to follow doctors' instructions. These are fruitful areas for future research with significant implications for development policy.

7 References

Acemoglu, D., Angrist, J. (2000). *"How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws"*. NBER Macroeconomics Annual, 1 January 2000, Vol.15, pp.9-59

Angrist, J., Pischke, J. (2010). *"Mostly Harmless Econometrics: An Empiricist's Companion"*. Princeton University Press

Anh, D., Tacoli, C. and Thanh, H. (2003). *"Migration in Vietnam: A Review of Information On Current Trends and Patterns, and Their Policy Implications"*. Paper presented at the Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia, 22–24 June 2003, Dhaka, Bangladesh

Banister, J. (1985). *"The Population of Vietnam"*. United States, Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census

Benjamini, Y., Krieger, A., Yekutieli, D. (2006). *"Adaptive Linear Step-up Procedures That Control the False Discovery Rate"*. Biometrika, 1 September 2006, Vol.93(3), pp.491-507

Bingley, P., Christensen, K., and Jensen, V. (2009). *"Parental Schooling and Child Development: Learning from Twin Parents"*. Danish National Centre for Social Research Social Policy and Welfare Services Working Paper 07:2009

Björklund, A., Lindahl, M., Plug, E. (2006). *"The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data"*. The Quarterly Journal of Economics, 1 August 2006, Vol.121(3), pp.999-1028

Borghans, L., Duckworth, A., Heckman, J., Weel, B. (2008). *"The Economics and Psychology of Personality Traits"*. The Journal of Human Resources, 1 October 2008, Vol.43(4), pp.972-1059

Bound, J., Jaeger, D., Baker, R. (1995). *"Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak"*. Journal of the American Statistical Association, 1995, Vol.90(430), p.443-450

Bowles, S., Gintis, H., Osborne, M. (2001). *"The Determinants of Earnings: A Behavioral Approach"*. Journal of Economic Literature 39 (4), 1137-1176

Bowles, S., & Gintis, H. (1976). *"Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life"*. New York, NY: Basic Books

Brierova, L., and Duflo, E. (2004) *"The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?"*. NBER Working Paper No. 10513

Caldwell, J. (1979). *"Education as a Factor in Mortality Decline: An Examination of Nigerian Data"*. Population Studies, 1 November 1979, Vol.33(3), pp.395-413

Capron C., Duyme, M. (1989). *"Assessment of Effects of Socio-economic Status on IQ in a Full Cross-fostering Study"*. Nature, 1989, Vol.340(6234), p.552

Carneiro, P., Meghir, C., Parey, M. (2013). *"Maternal Education, Home Environments, and the Development of Children and Adolescents"*. Journal of the European Economic Association, 2013, Vol.11, pp.123-160

Chandra, R. (2002). *"Nutrition and the Immune System from Birth to Old Age"*. European Journal of Clinical Nutrition, 2002, Vol.56(S3), p.S73

Cheung, V., Bocking, A., Dasilva, O. (1995). *"Preterm Discordant Twins: What Birth Weight Difference is Significant?"*. American Journal of Obstetrics and Gynecology, 1995, Vol.172(3), pp.955-959

Coneus, K., Laucht, M., Reuß, K. (2012). *"The Role of Parental Investments for Cognitive and Noncognitive Skill Formation: Evidence for the First 11 Years of Life"*. Economics and Human Biology, March 2012, Vol.10(2), pp.189-209

Cunha, F., Heckman, J., Schennach, S. (2010). *"Estimating the Technology of Cognitive and Noncognitive Skill Formation"*. Econometrica 78 (3), 883-931

Davis-Kean, P. (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, June 2005, Vol.19(2), pp.294-304

Dercon, S. and Krishnan, P. (2009). *"Poverty and the Psycho-Social Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries"*. Children, Youth and Environments 19(2): 138-163.

Dercon, S., Sánchez, A. (2013). *"Height in Mid Childhood and Psychosocial Competencies in Late Childhood: Evidence from Four Developing Countries"*. Economics and Human Biology, December 2013, Vol.11(4), pp.426-432

Dercon, S. and Singh, A. (2013). *"From Nutrition to Aspirations and Self-Efficacy: Gender Bias over Time among Children in Four Countries"*. World Development 45, 31–50

Dryler, H (1998). *"Parental Role Models, Gender and Educational Choice"*. The British Journal of Sociology, 1 September 1998, Vol.49(3), pp.375-398

Duflo, E. (2001). *"Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment"*. American Economic Review, September 2001, Vol.91(4), pp.795-813

Ermisch, J. (2008). *"Origins of Social Immobility and Inequality: Parenting and Early Childhood Development"*. National Institute Economic Review, 205(1), 62-71

Government of India. (2013). *"National Early Childhood Care and Education Policy"*. Ministry of Women and Child Development, Government of India, 27 September 2013

Güneş, P. (2015). *"The Role of Maternal Education in Child Health: Evidence from a Compulsory Schooling Law"*. Economics of Education Review, August 2015, Vol.47, pp.1-16

Hægeland, T., Kirkebøen, L., Raaum, O., and Salvanes, K. (2010). *"Why Children of College Graduates Outperform their Schoolmates: A Study of Cousins and Adoptees"*. Working Paper, Norwegian School of Economics

Hardy, A. (2000). *"Agricultural Colonisation in Lam Dong Province, Vietnam"*. Special Issue on Migration, Markets and Social Change in the Highlands of Vietnam, Asia Pacific Viewpoint 41(1): 23–34.

Hausman, J. (1978). *"Specification Tests in Econometrics"*. Econometrica, 1 November 1978, Vol.46(6), pp.1251-1271

Heckman, J. (2007). *"The Economics, Technology, and Neuroscience of Human Capability Formation"*. NBER Working Paper No. 13195

Heckman, J. and Rubinstein, Y. (2001). *"The Importance of Noncognitive Skills: Lessons from the GED Testing Program"*. American Economic Review 91(2):145-9

Holmlund, H., Lindahl, M., Plug, E. (2011). *"The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods"*. Journal of Economic Literature, 2011, Vol.49(3), pp.615-651

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G., Williams, J. (1979). *"Who Gets Ahead? The Determinants of Economic Success in America"*. New York, NY: Basic Books

Li, H., Luo, Y. (2004). *"Reporting Errors, Ability Heterogeneity, and Returns to Schooling in China"*. Pacific Economic Review, 2004, Vol.9(3), pp.191-207

Loehlin, J., Horn, J., Willerman, L. (1989). *"Modeling IQ Change: Evidence from the Texas Adoption Project"*. Child Development, 1 August 1989, Vol.60(4), pp.993-1004

Marr, D., White, C. (1988). *"Postwar Vietnam: Dilemmas in Socialist Development"*. Ithaca, N.Y : Southeast Asia Program, Cornell University

Masten, A., Shaffer, A. (2006). *"How Families Matter in Child Development: Reflections from Research on Risk and Resilience"*. In: Clarke-Stewart A, Dunn J (eds) Families Count: Effects on Child and Adolescent Development. Cambridge University Press, New York, pp 5-25

Maurin, E., Mcnally, S. (2008). *"Vive la Révolution! Long - Term Educational Returns of 1968 to the Angry Students"*. Journal of Labor Economics, 2008, Vol.26(1), pp.1-33

Neiss, M., Rowe, D. (2000). *"Parental Education and Child's Verbal IQ in Adoptive and Biological Families in the National Longitudinal Study of Adolescent Health"*. Behavior Genetics, 2000, Vol.30(6), pp.487-495

Nguyen, N. (2008). *"An Assessment of the Young Lives Sampling Approach in Vietnam"*. Young Lives (Project), DFID (University of Oxford)

Oreopoulos, P., Page, M., Stevens, A. (2006). *"The Intergenerational Effects of Compulsory Schooling"*. Journal of Labor Economics October 2006, Vol.24(4), pp.729-760

Pfeffer, F (2008). *"Persistent Inequality in Educational Attainment and its Institutional Context"*. European Sociological Review, 2008, Vol. 24(5), pp.543-565

Reichman, N., Pagnini, D. (1997). *"Maternal Age and Birth Outcomes: Data from New Jersey"*. Family Planning Perspectives, 1 November 1997, Vol.29(6), pp.268-295

Sacerdote, B. (2000). *"The Nature and Nurture of Economic Outcomes"*. NBER Working Paper No. 7949

Sacerdote, B. (2007). *"How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees"*. The Quarterly Journal of Economics, 1 February 2007, Vol.122(1), pp.119-157

Scarr, S., Weinberg, R. (1977). *"Intellectual Similarities within Families of both Adopted and Biological Children"*. Intelligence, 1977, Vol.1(2), pp.170-191

Strauss, J., Thomas, D. (2007). *"Health over the Life Course"*. Chapter 54, Handbook of Development Economics, 2007, Vol.4, pp.3375-3474

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2004). *"The Effective Provision of Pre-school Education (EPPE). Project findings for Pre-school to end of Key Stage 1"*. London: Department for Education and Science

United Nations. (2014). *"Young Children as a Basis for Sustainable Development"*. Issue Brief, Thematic Group on Early Childhood Development, Education and Transition to Work, 18 February 2014.

USAID. (2015). *"Multi-Sectoral Nutrition Strategy 2014 – 2025"*. USAID Technical Guidance Brief, May 2014

Vikram, K., Vanneman, R., Desai, S. (2012). *"Linkages between Maternal Education and Childhood Immunization in India"*. Social Science & Medicine, July 2012, Vol.75(2), pp.331-339

Young Lives. (2011). *"Young Lives Methods Guide: Sampling"*. Oxford: Young Lives, December 2001

Young Lives. (2013). *"Young Lives Survey Design and Sampling in Viet Nam"*. Oxford: Young Lives, December 2014

8 Tables

Table 3: Summary Statistics for the Younger Cohort of Young Lives Vietnam (Young Lives, 2013)

Table 3			Number	%
Gender		Male	992	51.5
		Female	936	48.5
Location		Urban	392	20.4
		Rural	1529	79.6
Gender of household head		Male	1630	85.6
		Female	274	14.4
Caregiver's education		None	4.5	10.8
		0-4 years	207	14.8
		5-8 years	282	35.0
		9 years plus	669	39.4
Average household size			753	
Ethnicity group		Majority Kinh	1652	85.7
		Minority ethnic group	276	14.3
Region	North	Northern Uplands	19.9	19.9
		Red River Delta	20.3	20.3
	South	Central Coastal urban	18.9	18.9
		Central coastal rural	20.5	20.5
		Mekong River Delta	20.4	20.4

Table 4: Relationship between Parental Education and Maths Ability

Table 4	Round 3 Maths Score (Rasch score)					Round 4 Maths Score (percentile score)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
OLS	2.45	2.45	2.42	2.56	2.00	2.62	2.62	2.53	2.59	1.93
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
2SLS	2.60	2.59	2.93	3.16	2.60	2.80	2.80	3.38	3.43	2.96
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.004)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.02)**
LIML	2.62	2.61	2.97	3.21	2.72	2.83	2.83	3.42	3.46	3.08
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)**	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)**
Observations										
OLS	746	746	746	746	746	734	734	734	734	734
IV	745	745	745	745	745	733	733	733	733	733
Controls										
Individual	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Household	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Community	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Underidentification test	59.98	59.81	41.72	33.17	21.14	54.07	54.33	42.30	33.67	18.96
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification test (bias)	16.76	16.72	11.31	8.72	5.12	14.91	15.00	11.17	8.69	4.57
	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(<30%)	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(>30%)
Overidentification test	5.78	5.75	3.26	2.61	3.42	7.61	7.53	1.56	1.03	1.58
	(0.12)	(0.12)	(0.20)	(0.27)	(0.18)	(0.06)**	(0.06)**	(0.46)	(0.60)	(0.46)
Endogeneity test	0.13	0.108	1.03	1.01	0.66	0.14	0.14	2.04	1.63	1.23
	(0.72)	(0.74)	(0.60)	(0.61)	(0.72)	(0.71)	(0.71)	(0.36)	(0.44)	(0.54)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

TABLE 5: RELATIONSHIP BETWEEN PARENTAL EDUCATION AND VERBAL ABILITY

Table 5	**Round 3 Vocab Score (Rasch score)**					**Round 4 Vocab Score (percentile score)**				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
OLS	2.23	2.23	2.17	2.14	2.00	1.78	1.78	1.64	1.60	1.72
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
2SLS	2.79	2.78	3.07	3.15	3.39	1.47	1.47	2.06	2.13	2.61
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
LIML	2.88	2.87	3.11	3.19	3.53	1.20	1.21	2.06	2.13	2.64
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.12)	(0.11)	(0.00)***	(0.00)***	(0.00)***
Observations										
OLS	732	732	732	732	732	756	756	756	756	756
IV	731	731	731	731	731	755	755	755	755	755
Controls										
Individual	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Household	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Community	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Underidentification test	61.89	62.20	49.22	38.99	25.01	51.74	51.88	40.81	32.39	18.71
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification test (bias)	17.53	17.67	13.77	10.60	6.33	14.41	14.47	11.43	8.72	4.64
	(<5%)**	(<5%)**	(<5%)**	(<10%)*	(<20%)	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(>30%)
Overidentification test	9.30	8.80	2.06	1.83	2.40	19.77	19.50	0.28	0.39	0.40
	(0.03)**	(0.03)**	(0.36)	(0.40)	(0.30)	(0.00)***	(0.00)***	(0.87)	(0.82)	(0.82)
Endogeneity test	2.59	2.50	6.92	6.18	5.57	0.03	0.03	3.13	3.34	3.22
	(0.11)	(0.11)	(0.03)**	(0.05)**	(0.06)*	(0.86)	(0.86)	(0.21)	(0.19)	(0.20)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

TABLE 6: RELATIONSHIP BETWEEN PARENTAL EDUCATION AND READING ABILITY

Table 6	**Round 3 Reading Score (Rasch score)**					**Round 4 Reading Score (percentile score)**				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
OLS	2.43	2.43	2.27	2.26	2.03	2.48	2.48	2.38	2.46	1.93
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
2SLS	1.64	1.62	2.16	2.06	1.65	2.31	2.32	3.09	3.24	2.96
	(0.00)***	(0.004***	(0.00)***	(0.00)***	(0.10)*	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)**
LIML	1.56	1.55	2.16	2.05	1.64	2.28	2.29	3.10	3.25	3.00
	(0.01)**	(0.01)**	(0.00)***	(0.00)***	(0.11)	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.02)**
Observations										
OLS	734	734	734	734	734	725	725	725	725	766
IV	733	733	733	733	733	724	724	724	724	765
Controls										
Individual	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Household	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Community	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Underidentification test	58.98	58.80	40.95	33.00	21.95	54.35	54.60	41.50	33.07	19.83
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification test (bias)	16.43	16.38	11.17	8.70	5.35	14.98	15.08	10.95	8.54	4.78
	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(<30%)	(<10%)*	(<10%)*	(<10%)*	(<10%)*	(<30%)
Overidentification test	4.79	4.61	0.39	0.43	0.42	9.80	9.86	0.61	0.57	0.64
	(0.19)	(0.20)	(0.82)	(0.81)	(0.81)	(0.02)**	(0.02)**	(0.74)	(0.75)	(0.72)
Endogeneity test	1.41	1.47	2.23	2.11	2.11	0.05	0.04	2.65	2.24	2.45
	(0.24)	(0.23)	(0.33)	(0.35)	(0.35)	(0.83)	(0.84)	(0.27)	(0.33)	(0.29)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

Table 7: Relationship between Parental Education and Self-Efficacy

Table 7	**Round 3 Self-efficacy**					**Round 4 Self-efficacy**				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
OLS	0.022	0.022	0.020	0.020	0.022	0.025	0.025	0.024	0.022	0.032
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
2SLS	0.002	0.002	0.007	0.005	(0.006)	0.022	0.022	0.022	0.023	0.032
	(0.91)	(0.93)	(0.74)	(0.82)	(0.88)	(0.17)	(0.17)	(0.20)	(0.26)	(0.30)
LIML	0.002	0.001	0.006	0.005	(0.008)	0.022	0.022	0.022	0.023	0.032
	(0.93)	(0.95)	(0.77)	(0.85)	(0.85)	(0.17)	(0.18)	(0.20)	(0.26)	(0.30)
Observations										
OLS	772	772	772	772	772	769	769	769	769	769
IV	771	771	771	771	771	768	768	768	768	768
Controls										
Individual	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Household	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Community	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Underidentification	50.73	50.73	36.91	29.62	17.14	52.11	52.04	41.12	33.04	19.03
test	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification	13.61	13.61	9.76	7.56	4.08	14.27	14.25	11.37	8.75	4.63
test (bias)	(<10%)*	(<10%)*	(<10%)*	(<10%)*	(>30%)	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(>30%)
Overidentification	1.53	1.43	1.43	1.72	1.22	0.03	0.03	0.01	0.10	0.04
test	(0.68)	(0.70)	(0.49)	(0.42)	(0.54)	(1.00)	(1.00)	(0.99)	(0.95)	(0.98)
Endogeneity test	1.52	1.56	0.84	0.68	0.98	0.02	0.03	0.33	0.30	0.41
	(0.22)	(0.21)	(0.66)	(0.71)	(0.61)	(0.88)	(0.87)	(0.85)	(0.86)	(0.82)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

Table 8: Relationship between Parental Education and Self-Esteem

Table 8	**Round 3 Self-esteem**					**Round 4 Self-esteem**				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
OLS	0.009	0.009	0.008	0.013	0.026	0.009	0.009	0.008	0.008	0.020
	(0.09)*	(0.09)*	(0.13)	(0.02)**	(0.00)***	(0.03)**	(0.03)**	(0.09)*	(0.07)*	(0.00)***
2SLS	(0.006)	(0.006)	0.004	0.013	0.035	(0.000)	(0.000)	0.001	0.013	0.040
	(0.75)	(0.77)	(0.84)	(0.57)	(0.34)	(0.98)	(0.99)	(0.54)	(0.46)	(0.16)
LIML	(0.008)	(0.007)	0.004	0.014	0.037	(0.002)	(0.002)	0.001	0.014	0.043
	(0.71)	(0.74)	(0.85)	(0.59)	(0.38)	(0.90)	(0.91)	(0.55)	(0.47)	(0.18)
Observations										
OLS	778	778	778	778	778	769	769	769	769	769
IV	777	777	777	777	777	768	768	768	768	768
Controls										
Individual	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓
Household	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Community	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Underidentification	52.22	52.15	38.84	31.53	18.97	52.11	52.04	41.12	33.04	19.03
test	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification	14.38	14.36	10.99	8.52	4.71	14.27	14.25	11.37	8.75	4.63
test (bias)	(<10%)*	(<10%)*	(<10%)*	(<10%)*	(<30%)	(<10%)*	(<10%)*	(<5%)**	(<10%)*	(>30%)
Overidentification	5.23	5.16	2.96	2.90	2.61	7.37	7.40	3.46	3.24	2.55
test	(0.16)	(0.16)	(0.23)	(0.23)	(0.27)	(0.06)*	(0.06)*	(0.18)	(0.20)	(0.28)
Endogeneity test	0.60	0.56	4.52	3.88	4.12	0.47	0.46	2.51	2.40	2.44
	(0.44)	(0.46)	(0.10)	(0.14)	(0.13)	(0.49)	(0.50)	(0.29)	(0.30)	(0.30)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

Table 9: Relationship between Parental Education and Aspirations

Table 9		**Round 4 aspirations**				
		(1)	(2)	(3)	(4)	(5)
OLS		0.20	0.20	0.19	0.19	0.18
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
2SLS		0.19	0.19	0.25	0.26	0.31
		(0.01)***	(0.01)***	(0.00)***	(0.00)***	(0.02)**
LIML		0.19	0.19	0.26	0.26	0.32
		(0.03)**	(0.03)**	(0.00)***	(0.00)**	(0.03)**
Observations	OLS	761	761	761	761	761
	IV	760	760	760	760	760
Controls	Individual	✗	✓	✓	✓	✓
	Household	✗	✗	✓	✓	✓
	Community	✗	✗	✗	✓	✗
	Region FE	✗	✗	✗	✗	✓
Underidentification test		52.67	52.64	40.28	32.44	19.10
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Weak identification test (bias)		14.49	14.48	11.13	8.63	4.66
		(<10%)*	(<10%)*	(<5%)**	(<10%)*	(>30%)
Overidentification test		7.93	7.93	1.80	1.47	1.63
		(0.05)**	(0.05)**	(0.41)	(0.48)	(0.44)
Endogeneity test		0.06	0.07	2.37	2.51	2.47
		(0.81)	(0.79)	(0.31)	(0.29)	(0.29)

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

Table 10: Results of First Stage of IV

Table 10	Average parental education					Number of siblings			
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(5)
Avg aggregate parental education in the South	0.91 (0.00)***	0.91 (0.00)***	0.92 (0.00)***	0.79 (0.00)***	0.51 (0.011)**	0.19 (0.00)***	0.19 (0.00)***	0.20 (0.00)***	0.23 (0.00)***
Both parents fully exposed to reforms	2.01 (0.01)**	2.02 (0.01)**	2.03 (0.01)**	1.82 (0.02)**	1.17 (0.08)*	(1.93) (0.00)***	(1.92) (0.00)***	(1.89) (0.00)***	(1.83) (0.00)***
One parent fully exposed to reforms	1.85 (0.00)***	1.86 (0.00)***	1.90 (0.00)***	1.66 (0.00)***	1.00 (0.01)***	(1.25) (0.00)***	(1.21) (0.00)***	(1.17) (0.00)***	(1.11) (0.00)***
Aggregate potential years of exposure to reforms	0.20 (0.02)**	0.19 (0.02)**	0.19 (0.03)**	0.16 (0.05)**	0.13 (0.05)**	(0.16) (0.00)***	(0.16) (0.00)***	(0.16) (0.00)***	(0.16) (0.00)***
Observations	786	786	786	786	786	786	786	786	786
Individual controls	✗	✓	✓	✓	✓	✓	✓	✓	✓
Household controls	✗	✗	✓	✓	✓	✗	✓	✓	✓
Community controls	✗	✗	✗	✓	✗	✗	✗	✓	✗
Region FE	✗	✗	✗	✗	✓	✗	✗	✗	✓

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions

Table 11: Relationship between Mother's Age and Premature Birth

Table 11	**Premature birth**	
	(1)	(2)
Mother's age	0.00	(0.00)
	(0.49)	(0.90)
Observations	1960	1870
Controls	✗	✓

Table 12: Relationship between Migrant Status and Parental Education

Table 12	**Parental Education**	
	(1)	(2)
Migrant (1=Migrant,	(0.15)	(0.12)
0=Non-migrant)	(0.59)	(0.58)
Observations	787	787
Controls	✗	✓

Table 13: Results for the Test for Common Pre and Post Trends

Table 13	Difference in education	
	Pre-1981	Post-1981
Year of school entry	(0.16)	0.02
	(0.08)*	(0.83)
Observations	13	8

Table 14: Test for Robustness of Instrument using Placebo Reforms

Table 14	**1976 placebo reform**			**1986 placebo reform**		
	Average parental education		Number of siblings	Average parental education		Number of siblings
	(1)	(2)	(1)	(1)	(2)	(1)
Average aggregate parental education in the South	0.92	0.51	0.37	0.83	0.49	(0.07)
	(0.00)***	(0.00)***	(0.00)***	(0.01)***	(0.08)*	(0.43)
Both parents fully exposed to reforms	3.18	1.97	(2.02)	(0.31)	0.28	(2.38)
	(0.00)***	(0.02)**	(0.00)***	(0.790)	(0.793)	(0.00)***
One parent fully exposed to reforms	1.31	1.19	(1.22)	1.12	0.34	(1.35)
	(0.06)*	(0.02)**	(0.00)***	(0.05)**	(0.53)	(0.00)***
Aggregate potential years of exposure to reforms	0.23	0.13	(0.11)	0.14	0.08	(0.21)
	(0.11)	(0.21)	(0.03)**	(0.10)*	(0.26)	(0.00)***
Observations	786	786	786	786	786	786
Full controls included?	✗	✓	✓	✗	✓	✓

Note: Two-sided test p-values are given in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Robust standard errors used in all regressions