

## Early Childhood Education and Cognitive Outcomes in Adolescence: A Longitudinal Study from Vietnam

Robert Rogers<sup>1</sup>  
Doan Hai Ma<sup>1</sup>  
Tra Nguyen<sup>1</sup>  
Ngoc Anh Nguyen<sup>1,2</sup>

<sup>1</sup> Development and Policies Research Center (DEPOCEN), Hanoi, Vietnam

<sup>2</sup> Department of Research, IPAG Business School, Paris, France

<sup>3</sup> We use “Early Childhood Education” and “Preschool” interchangeably in this paper. Both terms indicate educational programs for children aged 3-6 years old.

The DEPOCEN WORKING PAPER SERIES disseminates research findings and promotes scholar exchanges in all branches of economic studies, with a special emphasis on Vietnam. The views and interpretations expressed in the paper are those of the author(s) and do not necessarily represent the views and policies of the DEPOCEN or its Management Board. The DEPOCEN does not guarantee the accuracy of findings, interpretations, and data associated with the paper, and accepts no responsibility whatsoever for any consequences of their use. The author(s) remains the copyright owner.

DEPOCEN WORKING PAPERS are available online at <http://www.depocenwp.org>

# Early Childhood Education and Cognitive Outcomes in Adolescence: A Longitudinal Study from Vietnam

*Robert Rogers<sup>1</sup>, Doan Hai Ma<sup>1</sup>, Tra Nguyen<sup>1</sup> and Ngoc Anh Nguyen<sup>1,2</sup>*

## **Abstract:**

Previous research shows that Early Childhood Education (ECE) positively impacts cognitive outcomes later in life. Few studies examine the impacts of time spent in ECE in developing countries. We use data from the Young Lives project in Vietnam with 2SLS regressions to estimate the impact of years spent in ECE on cognitive outcomes in adolescence. **We find that one extra year in ECE corresponds to 21.8 percentage point (1.25 SD) and 30.8 percentage point (2.78 SD) increases in math and verbal cognition scores, respectively.** Our estimates suggest that ECE is highly effective in Vietnam and is a potential strategy for bridging educational outcomes gaps.

Keywords: early childhood education; education outcomes; Vietnam; cognitive outcomes

## **1. Introduction**

Much research in recent years, such as that of James Heckman (2011), focuses on the advantages of investment in Early Childhood Education (ECE)<sup>3</sup>. Notably, greater investment in ECE has been shown to lead to higher test scores and better cognitive skills later in school (Ladd et al 2014), which in turn lead to increased educational success in secondary school and beyond (Sisco et al 2014). Heckman (2006) also finds that increased ECE investment also has a disproportionately positive impact on the educational and cognitive outcomes of disadvantaged children.

Improved academic and cognitive skills in adolescence, as one might expect, significantly affect future education success, labor market outcomes, wages, and other factors that lead to happier and more productive populations (Heckman et al 2006). Furthermore, by improving the outcomes of disadvantaged students and thereby increasing educational equity, ECE can lead to greater economic and labor force mobility and lower economic inequality (Magnuson and Duncan 2016; Corak 2013). Heckman et al (2010) have also shown that investment in ECE is economically cost-

---

1 Development and Policies Research Center, Hanoi, Vietnam

2 Department of Research, IPAG Business School, Paris, France

<sup>3</sup> We use “Early Childhood Education” and “Preschool” interchangeably in this paper. Both terms indicate educational programs for children aged 3-6 years old.

effective, with annual returns of 7-10% on the initial investment, well above those of private equity indices and estimates of other types of educational investment.

Longitudinal study of the effects of ECE programs has been well documented in cases such as the influential Perry Preschool Program, which tracked participants in a US-based ECE program well into adulthood. Schweinhart et al (2005) find that the participants in the Perry program had higher earnings and lower crime rates in adulthood, as well as higher chances of success in secondary school. Unfortunately, very little research examines the effects of ECE in the developing world, where rising inequality and development challenges make the question of efficacious education investment even more pertinent, as Burger (2010) highlights. In Vietnam, for example, only one small interventional study by Watanabe et al (2005) on ECE impacts has been conducted. Structural and labor force differences across countries can have significant effects on the predictors of education and labor force success, as Allmendinger (1989) finds, so this lack of formal study leaves developing country policymakers in the dark regarding the comparative advantage ECE investment could provide. In addition, most existing studies of ECE programs in developing countries focus on isolated interventional programs, which, while critically important, do not provide robust information of the effect ECE has on nationwide populations that is crucial for policy development and implementation (Nores and Barnett 2010).

Vietnam has great potential for the increased study of and investment in ECE. First of all, Vietnam experienced extremely rapid transition from lower-income to middle-income status, and is now striving to transition to a more developed and knowledge-based economy, highlighting the critical importance of education in this transitory period (see e.g. Ohno 2009). Second, Dang and Rogers (2013) find that Vietnamese society places a high value on educational success and investment generally, suggesting the willingness of both governments and families to commit further to ECE investment, as they have in the past (London 2006). In addition, as Benjamin et al (2016) explain, Vietnam continues to suffer from significant inequalities that center on the urban and rural poor and ethnic minority groups, which is often correlates with educational disparities. Finally, while Vietnam has made great strides in prioritizing and promoting ECE, Boyd and Phuong (2017) find that access remains limited for ethnic minorities and disadvantaged groups in rural and mountainous areas, indicating significant provision gaps and room for further development.

Our study thus employs longitudinal data from the Young Lives (YL) project in Vietnam to examine the effect of years spent in ECE programs from ages three to six on outcomes later in life. We specifically focus on outcomes in terms of verbal and quantitative reasoning at age 12. The YL data are very rich and provide us with the opportunity to consider a wide variety of factors measured at various ages throughout the children's development, allowing for a robust estimation of the effects of ECE in the study cohort. By estimating the marginal effects of extra years of ECE with the YL data, we are able to identify the impact that even marginally increased investment in such programs could have for Vietnam, where ECE participation rates are relatively high but significant disparities continue to exist along socioeconomic lines, as disadvantaged groups such as ethnic minorities and the rural poor are well represented in the YL cohort (see Pells 2011).

We structure the remainder of this paper as follows. We start by providing a brief overview of the literature on ECE and on the current state of ECE in Vietnam. We present YL data and descriptive

statistics in the following section, followed by a discussion of the estimation methods used and the presentation of our results and conclusions.

## **2. Background**

### ***2.1. Early Childhood Education***

Previous research has found that Early Childhood Education is effective at improving cognitive and academic outcomes later in life for a variety of reasons relating to educational development and equity. For one, Birch and Ladd (1997) find that developing teacher-student relationships before starting primary school can help with the psychological challenges associated with the commencement of formal education, thereby improving academic performance later on. Gorey (2001) finds that, at young ages, the minds of children are incredibly malleable, and the effects that the institutional and educational environments ECE programs provide have on future outcomes compounds accordingly, including both cognitive and socioemotional effects. In addition, as Heckman (2011) and others highlight, ECE can help students who would have started behind to be at the same level as their peers at the debut of primary school, contributing significantly to educational equity and improving outcomes for the most disadvantaged members of society. Finally, Phillips et al (1987) find that years spent in ECE programs affect math and reading achievement, and Currie (2001) finds that the years spent in ECE has broad effects on a variety of cognitive outcomes, suggesting that the time spent, not just simple participation, also has meaningful impacts on cognitive outcomes later in life.

Despite the proven benefits of participation in ECE programs, various structural and sociodemographic limitations exist that prevent full participation in ECE programs. In many cases, parents need to personally take their children to school. This requires time and a significant degree of parental involvement in the children's education, which has been shown to be a major factor in the success and retention rates of ECE programs (Miedel and Reynolds 2000; Vandenberg and Lazzari 2014). In addition, parents often need to pay for ECE, even in publicly subsidized systems. Some simply cannot afford this and others will be deterred by the cost, as discussed by Currie (2001) and others. As such, even with government subsidized ECE, these limitations may reduce participation in ECE, especially for rural populations in developing countries.

Despite a growing focus on the study of ECE around the world, a number of gaps exist in the literature. First of all, because of the highly context-dependent nature of education and education systems around the world, the external validity of past research is always somewhat questionable (Burger 2010; Jonsson and Svingby 2007). Without further investigation, we cannot be sure that well-known studies on the effects of ECE in the United States and the United Kingdom will generalize to developing countries like Vietnam. The effects of ECE and especially of time in ECE are not well studied in developing countries where the above-discussed limitations can be even more significant. Other than the small interventional study by Watanabe et al (2005), no longitudinal analysis of the impacts of ECE has been conducted in Vietnam. The Watanabe study found positive cognitive results of the examined ECE intervention, but focused solely on participation or non-participation in ECE, rather than the marginal effects of time spent in such programs. This ignores important differences between students who attend one class and students

who attend ECE programs for three years continuously. As such, further study is important to understand the impacts of ECE in countries like Vietnam, and to determine the marginal effects of extra time spent in ECE programs.

## ***2.2. The Early Childhood Education Landscape in Vietnam***

ECE in Vietnam has a history reaching back to the early 1900s, when ECE was provided to encourage women's participation in the largely agrarian economy (Boyd and Phuong 2017). However, given the history of war and conflict in the country, significant ECE reform was effectively impossible until reunification in 1975. The first nationwide education reforms took place in the early 1980s, and education has subsequently become a significant focus of the government, with Early Childhood Care and Education becoming an official department of the Ministry of Education in 1991. Since the department's creation, and well into the 2000s and 2010s, universal access to ECE has been a key priority and significant improvements have been made (Thao and Boyd 2014).

Today, Vietnam has a system of semi-public nursery schools (ages one to three) and preschools (ages three to six) that receive partial state support, and a growing number of fully public preschools that operate primarily in economically disadvantaged areas. Vietnam has also developed certain legal foundation to regulate the provision of ECE as well as contents of ECE curriculum. **Law of Education 2005 which was later updated into the Law of Education 2019 has regulations about the contents and methodologies for preschool education as well as the requirements for the preschool education curriculum. In the context of our paper, the regulation on General Preschool Education Curriculum No.17/2009/TT-BGDĐT issued by the Ministry of Education and Training in 2009 also regulates four sets of targets for preschool education, including physical development, cognitive development, lingual development and emotional and social skill development, as well as more detailed curriculum for each age from 3 to 6 years old.** Boyd and Phuong (2017) find that over half of the total population of children aged three to six (4.8 million children), were enrolled in preschools in 2015. However, access to ECE amongst ethnic minority groups and other disadvantaged groups, such as those living in remote areas, remains a pressing issue. ECE in Vietnam is also not compulsory and many programs require parents to pay tuition fees. As such, despite great advancements in ECE access due to a strong commitment by the government of Vietnam, further research is warranted to better understand the impacts of ECE in the country, and to tailor ECE policies and programs accordingly.

## **3. Estimation methods**

We aim to estimate the impact of time spent in preschool education on outcomes later in life. However, because the data do not extend beyond early adolescence for the YL Vietnam children, we focus on verbal cognition and math outcomes in early adolescence. Following the literature on cognitive outcomes and educational attainment, we specify cognitive test score outcomes as a function of a number of control covariates that capture family background (wealth index), parental education, geographical location, and other sociodemographic factors, as these variables have been used in studies such as Woldehanna (2012) and Le (2012) in the context of educational outcomes and the YL data. We specify our regression model as follows.

$$Test\ Score_t = \beta_1 + \beta_2 \times Pre\_ScL_t + \beta_3 \times X_t + u_t$$

Where  $TestScore_t$  is the observed results on cognitive test scores at round 4,  $Pre\_ScL_t$  represents the observed years spent in ECE at round 2, and  $X_t$  represents the set of control variables observed at various rounds.

However, the issue with this approach is that, while some variables included in the model are reliably exogenous, time spent in preschool is almost certainly endogenous to test score results and reflects the parents' role in the decision to enroll their child in non-mandatory ECE programs. The endogeneity issue with the main preschool variable thus obscures the results of the OLS model specified above.

Previous work in educational outcomes research, such as in Todd and Wolpin (2004), identifies two types of inputs that determine cognitive and academic achievement: endogenous inputs that in part reflect choices at the household level by parents, and exogenous inputs unaffected by parental preferences and decisions. As such, the primary issue with the OLS regression approach is that it does not sufficiently account for the effect of parental decision-making factors on the child's participation in ECE programs. We thus cannot reliably isolate our estimate of the causal impact of ECE on test scores in adolescence with a simple OLS regression.

The ideal way to address the endogeneity issue would be to conduct a randomized controlled trial that accounts for confounding factors and omitted variables to maximally isolate the effect of preschool on the outcome variables. Examples of this type of study abound in the literature. Notable recent examples include the longitudinal evaluation of the Abbott preschool program by Barnett, W. S. et al. (2013) and the Jamaica-based study on the labor market returns of an ECE intervention by Gertler P. et al. (2014). Both studies utilize a longitudinal approach with randomly assigned treatment and control groups to evaluate the impacts of ECE interventions.

A randomized longitudinal experiment for the purposes of the present study, however, would be incredibly costly, take many years to conduct, and potentially require the ethically questionable random assignment of individuals to control groups where they would not be able to benefit from ECE programs (Lilford and Jackson 1995; Cook and Payne 2002). As such, we pursue more feasible quasi-experimental approaches options with the YL data.

We employ a Two Stage Least Squares (2SLS) instrumental variable approach to help account for the selection bias driven by parental factors that make the preschool variable endogenous with test score outcomes. Instrumental variable regressions are often used in educational outcomes research to address endogeneity issues. For example, in perhaps the most similar published study to the present analysis, Cueto et al (2016) use the data from Peru and a 2SLS approach to estimate the impact of nutrition and participation in two different types of preschools on cognitive vocabulary and non-cognitive outcomes at age 8. This study employs mother's height to instrument for the child's nutrition and district-level preschool availability and district-level parental education levels as instruments, which they found to be both statistically valid and relevant. We therefore employ similar district- and household-level instruments in our 2SLS models.

The first stage of a 2SLS estimation regresses the endogenous variable on the exogenous covariates and instruments included in the model. We specify the first-stage equation as follows.

$$Pre\_ScL_t = \gamma_1 + \gamma_2 \times Capital\_District\_Distance_t + \gamma_3 \times X_t + \varepsilon_t$$

Where the observed preschool variable,  $Pre\_ScL_t$  is regressed on the instrument  $Capital\_District\_Distance_t$  as well as the set of control variables,  $X_t$ .

In the second stage, we use the predicted values from the first-stage estimations to regress the dependent outcome variables on the models' covariates, with the exception that we use the first-stage parameter estimates for the preschool variable rather than the observed measure of time spent in preschool.

$$Test\ Score_i = b_1 + b_2 \times \widehat{Pre\_ScL}_i + b_3 \times X_i + u_i$$

As we can see, this equation is identical to that of our OLS model above except that  $\widehat{Pre\_ScL}_i$ , which represents the predicted value for years in preschool from the first stage regression, replaces the observed value for the preschool variable.

The validity of instrumental variable approaches depends upon the relevance, validity and exogeneity of the instruments. In other words, instruments must correlate with the endogenous variable in the first stage and not correlate with the error term and the dependent variables in the second stage.

The instruments we employ are a commune-level variable for distance from home to the district capital, measured by time travelling to it. Distance is often used as an instrument for geographical barriers to participation or compliance with non-mandatory educational programs, particularly in developing countries, such as in Zhao and Glewwe (2010) and Falch et al (2013). We suspect that children in families living further away from district centers are less likely to attend ECE programs, which are non-compulsory in the case of Vietnam and require parents to drop their children off at school.

Regarding to the exclusion restriction, we are aware of a remote possibility that the distance to the district capitals can affect a child's cognitive and non-cognitive performance in adolescence through other channels rather than duration of preschool education. However, we expect that preschool is the primary medium, especially given that there was a period of 8 years between round 2—when the distance to the district capitals and preschool duration were recorded and round 4 when the child's performance was evaluated.

Finally, to further test the validity and relevance of our instruments, we tested the strength of the correlation between the instrument and the endogenous preschool variable in the first stage and found the correlation to be highly significant across both models ( $p < 0.01$ ). We perform Durbin and Wu-Hausman test of endogeneity test to ensure that the preschool variable was in fact endogenous with test score results, and to check the coherency of the instruments' effects.

Regarding to the weak instrument test, the F-statistics for both 2SLS models are also higher than 10 (14.2 for both models of verbal and quantitative reasoning).<sup>4</sup>

## 4. Data

### 4.1. *Young Lives*

The YL data come from a longitudinal study funded by the UK Department for International Development and conducted by the University of Oxford, which follows 12000 young people in four developing countries: Peru, India, Vietnam, and Ethiopia. In Vietnam, the Department of International Development at the University of Oxford, the Vietnam General Statistics Office, the Vietnam-based Center for Analysis and Forecasting, and the Vietnam Academy of Social Sciences conduct the design and fieldwork for all components of the YL study, in 31 communes from five provinces in Vietnam: Lao Cai (Northeast region), Hung Yen (Red River Delta), Da Nang (central urban center), Phu Yen (south-central coast) and Ben Tre (Mekong River Delta).

The YL project follows two cohorts in each of the four countries, an older cohort (n=1000) and a younger cohort (n=2000); our study uses the younger cohort panel data from Vietnam. **The younger cohort data come from five data collection rounds over the course of 15 years: round one in 2002 (6-18 months old), round two in 2006 (5-6 years old), round three in 2009 (8-9 years old), and round four in 2013 (12-13 years old) and round five in 2016 (15 years old).** Our study specifically uses the younger cohort's ECE data from round two, various other data from rounds one through four for control and instrumental variables, and verbal and quantitative reasoning outcomes data from round four.<sup>5</sup> All components of the YL study have undergone comprehensive ethical review by the University of Oxford, the Vietnam Union of Science and Technology Associations (VUSTA 2001) and the Hanoi School of Public Health.

[Insert Figure 1 here]

Some minor attrition exists in the younger cohort from the first to fourth round (3.6%), but was likely too small to significantly affect the validity of the sample. Additionally, as participation in the study was entirely voluntary, some students chose not to respond to various elements of the study, and such observations were excluded from the relevant models. Full details of the YL sampling and data collection methods in Vietnam have been discussed elsewhere (Nguyen 2008; Young Lives 2014).

Previous work with the YL data in Vietnam examines the effect of primary schooling (Glewwe et al 2014), extra-curricular classes (Le 2012), and a wide variety of nutritional, sociodemographic and household characteristics on cognitive and academic performance at later examination rounds (see e.g. Dearden et al 2017). However, the long-term effects of years spent in preschool and the effects of ECE on cognitive outcomes in adolescence have not yet been studied with these data.

---

<sup>4</sup> See Appendix table 6 and 7 for further details on F-statistics, endogenous test and over-identification tests' statistics.

<sup>5</sup> See Table 3 for descriptive statistics on all variables included in the model.

#### **4.2. Preschool variable**

Preschool education is measured in round 2 of the younger cohort with parent-directed questions asking about the time spent in each of the preschools the YL child has attended. Table 1 presents the breakdown of the preschool variable by a variety of different socioeconomic factors. As we can see, more time spent in preschool seems to skew in the direction of Kinh ethnicity and higher wealth index. This is largely to be expected given the well documented and above-discussed disparities between socioeconomic and ethnic groups in Vietnam.

[insert Table 1 here]

#### **4.3. Outcomes variables**

This study uses two outcomes indicators: a verbal cognition test and a mathematics test. The verbal cognition test is a Vietnam-adapted Peabody Picture Vocabulary Test (PPVT), a test widely used as a measure of verbal cognitive achievement (Todd and Wolpin 2003; Cueto et al 2009; Schady 2011). The mathematics test consists of 34 age-appropriate mathematic and quantitative reasoning questions. These two dependent outcomes variables were converted into percentage correct scores for inclusion in the study models as dependent variables. We also use normalized variables to determine the effect size of our final 2SLS estimates. Individuals missing both of the outcomes variables were excluded from the analyses.

[Insert Table 2 here]

The *descriptive* statistics in Table 2 above show the mean test scores of students grouped by the number of years they spent in preschool programs. As we can see, there is a clear upward trend in test scores in relation to the years spent in preschool education. We see this trend for both the verbal and quantitative reasoning outcomes.

#### **4.4. Control Variables and Descriptive Statistics**

Our models include a series of variables to account for the effect of socioeconomic, geographic, and other factors on the cognitive outcome variables. These controls include age, sex, family wealth index, ethnicity (Kinh majority vs. ethnic minorities), nutritional status (height-for-age Z scores), regional dummy variables with the poorest region (Lao Cai) as baseline, **dummy variables indicating the education level of the child's parents and employment status of the children's mothers.**

A factor that could lead to significant variation in ECE outcomes is the quality of the instruction in the early childhood centers of the populations in this study. Both time spent in ECE and the quality of the instruction can be predictors of the effectiveness of the ECE programs themselves, and so ideally the analysis would control for variations in ECE quality. However, data pertaining to ECE quality from the YL round two surveys show that 98% of parents consider their children's school "reasonably good" to "excellent," making it very hard to detect quality differences or to examine their impacts on cognitive outcomes. As such, we do not explicitly control for school quality in the present analysis.

Table 3 presents descriptive statistics for all outcome, test, and control variables used in the present analyses. The average child spent 1.8 years in ECE programs and was around 12 years and two months old (146 months). Boys (52%) and girls (48%) were nearly equally represented, and 86% of the children were from the Kinh (Viet) ethnic majority group, with the remaining 14% from H'mong, Tay, Thai, and other ethnic minority groups. While about 20% of the children in the sample came from each of the Lao Cai and Ben Tre provinces in the fourth round, only about 1% had moved to other provinces. As discussed above, some children and parents did not respond to individual questions, leading to a fair amount of variation in the number of observations across the examined variables.

[Insert table 3 here]

These descriptive statistics collectively present a picture of the situation in Vietnam that is similar to the prior research discussed above. There is expected variation of preschool attendance along the lines of ethnicity and socioeconomic status, and the breakdown of the test scores by years spent in preschool indicates a trend for a positive correlation. The descriptive findings presented above thus provide a solid foundation for the statistical estimations that will follow.

## 5. Estimation Results and Statistics

In this section, we present both OLS regression estimates and the estimates of the 2SLS instrumental variable models. Because of the endogeneity issues discussed in the estimation methods section, only the 2SLS results should be considered the final outputs for this study; the OLS results are intended for cross comparison and to illustrate the research process.

### 5.1. OLS Regression Estimates

Table 4 shows the estimates of the OLS regression models with the inclusion of the control variables discussed above. We can see that years in preschool have insignificant effects on Math score and statistically significant effect at the level of 10% on the PPVT score. In addition, the coefficients are small, indicating an increase of less than one percentage point per year of preschool education. In these models, the coefficients for family wealth and Kinh ethnicity are by far the strongest and most significant predictors of cognitive results.

[Insert Table 4 here]

These OLS models cannot adjust for the endogeneity of the preschool test variable. As the 2SLS analysis does account for this issue, we consider the following IV estimates the primary results of this paper.

### 5.2. 2SLS Regression Estimates

Table 5 presents estimates from the final 2SLS regression models. These models include the same control variables as the OLS models above. Our 2SLS models indicate that years spent in preschool

is a significant predictor of higher scores on both tests. Notably, our estimates find that one extra year of preschool education leads to 30.8 ( $p < 0.01$ ) and 21.8 ( $p < 0.01$ ) percentage point increases in verbal cognition and math test scores, respectively. In term of standardized effect sizes, the results show that one year at preschool translates to increases of 2.78 standard deviations (SD) on verbal cognition tests and 1.25 SD on quantitative reasoning scores.<sup>6</sup> As we can see, the estimate for verbal cognition is larger than the estimate for quantitative reasoning. One possible explanation for the difference in the effects of preschool education on test scores is that the education focuses more on developing the verbal skills for the children. More specifically, the Law of Education 2005 (Article 23) states that the contents of the curriculum of preschool education would be designed to help children learn how to respect their parents, grandparents and other older people, which strongly related to the verbal skills, while did not set any targets for their mathematic skills. The General Preschool Education Curriculum in 2009<sup>7</sup>, which gave more specific teaching contents for preschool education, also followed this orientation. As stated in the curriculum, lingual development was considered as one of the main “pillars” in the target of the curriculum<sup>8</sup>, while mathematics skills was only one sub-target in the cognitive development and the curriculum was designed to introduce some elementary concepts.

[Insert Tables 5 here]

The significant differences between the estimates of the OLS and 2SLS models suggests the presence of heterogeneous treatment effects among groups, and points to the large local average treatment effect for compliance group, as theorized by Angrist et al (1994). The validity of the models is supported by the above-discussed endogeneity test results ( $p < 0.01$ ), the theoretical justification of the validity and relevance of the instrument, and the instruments’ strong correlation ( $p < 0.01$ ) with the preschool variable in the first stage and the F-statistics (larger than 10). As such, we maintain that the 2SLS estimates present a reasonably accurate picture of the effect of years in ECE on cognitive outcomes in the YL Vietnam cohort.

## 6. Conclusions and Policy Implications

While previous research shows that ECE can have significantly positive impacts on cognitive and non-cognitive outcomes later in life, preschool’s effects have not been studied extensively in developing countries, and never on a large scale in Vietnam. Our empirical findings suggest a highly significant impact of years spent in ECE on outcomes in adolescence. In particular, we estimate that one extra year in ECE leads to a nearly 30.8 percentage point increase (2.78 SD) in verbal cognition scores and a 21.8 percentage point increase (1.25 SD) in math test scores. There are also significant impacts of family wealth, region, ethnicity, and maternal education on cognitive outcomes.

We find that years spent in ECE is a major predictor of better cognitive outcomes in early adolescence, indicating the important role preschool education does and can play in the education landscape in Vietnam. Furthermore, our results suggest that the cognitive achievement gaps

---

<sup>6</sup> See table 9 for full results in SD

<sup>7</sup> Decree No. 17/2019/TT-BGDĐT by the Ministry of Training and Education, issued on 25 July 2009.

<sup>8</sup> The others were physical development, cognitive development and emotional and social skill development.

established at the end of preschool remain or even grow wider in the longer term in Vietnam. This finding may have serious implications for policy in Vietnam and other developing countries with similar education landscapes. Finally, the observed variation in cognitive outcomes along socioeconomic and ethnic lines suggests that more attention should be paid to understanding and reducing disparities in education provision and quality. Investing in preschool education in these more disadvantaged populations may be an effective way to overcome such issues.

Further research is warranted to examine the cost-effectiveness of ECE programs and policies in developing countries, as well as the relative effectiveness of various features of ECE programs and pedagogies. This will allow governments and local and international actors to better understand the specific impacts of ECE programs and the mechanisms that drive their success.

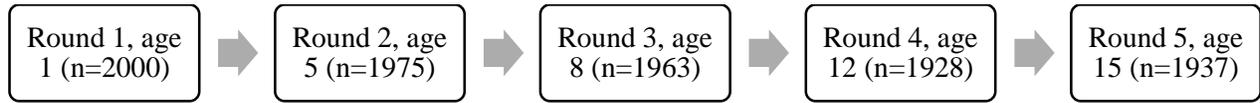
## References

- Allmendinger, J. 1989. "Educational systems and labor market outcomes." *European sociological review*, 231-250.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). "Identification of causal effects using instrumental variables". *Journal of the American statistical Association*, 91(434), 444-455.
- Benjamin, D., Brandt, L., & McCaig, B. 2016. "Growth with Equity: Income Inequality in Vietnam, 2002–14." Benjamin, Dwayne and Brandt, Loren and McCaig, Brian, Growth with Equity: Income Inequality in Vietnam, 2002–14. IZA Discussion Paper No. 10392.
- Birch, S. H., & Ladd, G. W. (1997). "The teacher-child relationship and children's early school adjustment." *Journal of school psychology*, 35(1), 61-79.
- Boyd, W., & Phuong, T. D. 2017. "Early childhood education in Vietnam: history and evaluation of its policies." In *Early Childhood Education Policies in Asia Pacific* (pp. 263-283). Springer Singapore.
- Burger, K. 2010. "How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds." *Early childhood research quarterly*, 25(2), 140-165.
- Cook, T. D., & Payne, M. R. 2002. "Objecting to the objections to using random assignment in educational research." *Evidence matters: Randomized trials in education research*, 150-178.
- Corak, M. 2013. "Income inequality, equality of opportunity, and intergenerational mobility." *The Journal of Economic Perspectives*, 27(3), 79-102.
- Cueto, S., Leon, J., Guerrero, G., & Muñoz, I. 2009. "Psychometric characteristics of cognitive development and achievement instruments in Round 2 of Young Lives." Young Lives Working Paper.
- Currie, Janet. "Early childhood education programs." *The Journal of Economic Perspectives* 15.2 (2001): 213-238.
- Dang, H. A., & Rogers, H. 2013. "The decision to invest in child quality over quantity: Household size and household investment in education in Vietnam." *World Bank, Development Research Group, Poverty and Inequality Team and Human Development Network, Education Unit*.
- Dearden, Kirk A., et al. 2017. "Does household access to improved water and sanitation in infancy and childhood predict better vocabulary test performance in Ethiopian, Indian, Peruvian and Vietnamese cohort studies?." *BMJ open* 7.3: e013201.
- Downey, D. B. (2001). "Number of siblings and intellectual development: The resource dilution explanation." *American Psychologist*, 56(6-7), 497.
- Elango, S., García, J. L., Heckman, J. J., & Hojman, A. 2015. "Early childhood education." National Bureau of Economic Research.
- Falch, T., Lujala, P., & Strøm, B. (2013). "Geographical constraints and educational attainment." *Regional Science and Urban Economics*, 43(1), 164-176.
- Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., ... & Grantham-McGregor, S. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." *Science*, 344(6187), 998-1001.
- Glewwe, P., Krutikova, S., & Rolleston, C. 2014. "Do Schools Reinforce or Reduce Learning Gaps between Advantaged and Disadvantaged Students?: Evidence from Vietnam and Peru."

- Gorey, K. M. 2001. "Early childhood education: A meta-analytic affirmation of the short-and long-term benefits of educational opportunity." *School Psychology Quarterly*, 16(1), 9.
- Heckman, J. J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science*, 312(5782), 1900-1902.
- Heckman, J. J. 2011. "The Economics of Inequality: The Value of Early Childhood Education." *American Educator*, 35(1), 31.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics*, 94(1), 114-128.
- Heckman, J. J., Stixrud, J., & Urzua, S. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics*, 24(3), 411-482.
- Heckman, J., Pinto, R., & Savelyev, P. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *The American economic review*, 103(6), 2052-2086.
- Jonsson, A., & Svingby, G. 2007. "The use of scoring rubrics: Reliability, validity and educational consequences." *Educational research review*, 2(2), 130-144.
- Ladd, H. F., Muschkin, C. G., & Dodge, K. A. 2014. "From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina." *Journal of Policy Analysis and Management*, 33(1), 162-187.
- Le, T. D. 2012. "Do Extra Classes Improve Cognitive Test Scores?: Evidence from Vietnam." *Young Lives Working Paper*.
- Lilford, R. J., & Jackson, J. 1995. "Equipoise and the ethics of randomization." *Journal of the Royal Society of Medicine*, 88(10), 552.
- London, J. D. 2006. "Vietnam: The political economy of education in a 'Socialist' periphery." *Asia Pacific Journal of Education*, 26(1), 1-20.
- Magnuson, K., & Duncan, G. J. 2016. "Can early childhood interventions decrease inequality of economic opportunity?" *RSF*.
- Miedel, W. T., & Reynolds, A. J. (2000). "Parent involvement in early intervention for disadvantaged children: Does it matter?" *Journal of School Psychology*, 37(4), 379-402.
- Nguyen, N. 2008. An assessment of the Young Lives sampling approach in Vietnam.
- Nores, M., & Barnett, W. S. 2010. "Benefits of early childhood interventions across the world:(Under) Investing in the very young." *Economics of education review*, 29(2), 271-282
- Ohno, K. 2009. "Avoiding the middle-income trap: renovating industrial policy formulation in Vietnam." *ASEAN Economic Bulletin*, 26(1), 25-43.
- Pells, K. 2011. "Poverty and gender inequalities: evidence from Young Lives." Young Lives, Department of International Development, University of Oxford.
- Phillips, D., McCartney, K., and Scarr, S. "Child-care quality and children's social development." *Developmental psychology* 23.4 (1987): 537.
- Schady, N. 2011. "Parents' education, mothers' vocabulary, and cognitive development in early childhood: Longitudinal evidence from Ecuador." *American Journal of public health*, 101(12), 2299-2307.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. 2005. "Lifetime effects: the High/Scope Perry Preschool study through age 40."
- Sisco, S., Gross, A. L., Shih, R. A., Sachs, B. C., Glymour, M. M., Bangen, K. J., ... & Manly, J. J. 2014. "The role of early-life educational quality and literacy in explaining racial

- disparities in cognition in late life.” *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, gbt133.
- Thao, D. P., & Boyd, W. A. 2014. “Renovating early childhood education pedagogy: A case study in Vietnam.” *International Journal of Early Years Education*, 22(2), 184-196.
- Todd, P. E., & Wolpin, K. I. 2003. “On the specification and estimation of the production function for cognitive achievement.” *The Economic Journal*, 113(485).
- Vandenbroeck, M., & Lazzari, A. 2014. “Accessibility of early childhood education and care: A state of affairs.” *European Early Childhood Education Research Journal*, 22(3), 327-335.
- VUSTA. 2001. “Re: Decision on implementing the child poverty research (Young Lives) in Vietnam, Lien Hiep Cac Hoi Khoa Hoc va Ky Thuat Viet Nam.”
- Watanabe, K., Flores, R., Fujiwara, J., & Tran, L. T. H. 2005. “Early childhood development interventions and cognitive development of young children in rural Vietnam.” *The Journal of nutrition*, 135(8), 1918-1925.
- Woldehanna, T. 2012. “The Effects of Pre-school Attendance on the Cognitive Development of Urban Children Aged 5 and 8 Years: Evidence from Ethiopia.” Young Lives Working Paper.
- Young Lives. 2014. “Young Lives Survey Design and Sampling in Vietnam.”
- Young Lives. 2018. “Survey Design and Sampling: Round 5 in Vietnam.”
- Zhao, M., & Glewwe, P. 2010. “What determines basic school attainment in developing countries? Evidence from rural China.” *Economics of Education Review*, 29(3), 451-460.

Figure 1: Young Lives Younger Cohort Data Collection Rounds



Source: Young Lives Vietnam rounds 1, 2, 3, 4, 5 datasets and Lives, Y (2014, 2018)

Table 1: Sociodemographic factors by time in preschool, mean (SD)

Time in Preschool	Age in months	Female	Wealth index <sup>a</sup>	Kinh <sup>b</sup>	Observations
Not going to school	143.23 (4.12)	0.45 (0.50)	0.53 (0.16)	0.67 (0.47)	132
Less than 1 year	145.77 (3.55)	0.48 (0.50)	0.58 (0.11)	0.85 (0.36)	401
From 1 to less than 2 years	146.09 (3.56)	0.50 (0.50)	0.62 (0.13)	0.84 (0.37)	497
More than 2 years	147.85 (2.96)	0.46 (0.50)	0.65 (0.11)	0.92 (0.27)	523
All children	146.36 (3.66)	0.48 (0.50)	0.61 (0.13)	0.86 (0.35)	1553

<sup>a</sup> Wealth index is a composite index of family wealth

<sup>b</sup> Kinh is the majority ethnic group in Vietnam

Table 2: Test Score Results<sup>a</sup> by Time Spent in Preschool, mean (SD)

Time in Preschool	Verbal Cognition (PPVT)	Math	Observations
Not going to school	71.74 (14.40)	38.66 (17.10)	132
Less than 1 year	75.27 (10.63)	44.01 (14.22)	401
From 1 to less than 2 years	76.52 (10.76)	48.01 (17.01)	497
More than 2 years	78.64 (8.85)	52.14 (16.91)	523
All children	76.50 (10.67)	47.57 (16.81)	1553

<sup>a</sup> Test scores represent the percent of correct responses

<sup>b</sup> PPVT is a vocabulary test widely used as an indicator of verbal cognitive development

Table 3: Descriptive Statistics for all Variables

	Mean	SD	Observations
Age of child in months	146.36	3.66	1553
Sex of members	0.48	0.50	1553
Wealth index	0.49	0.17	1553
Kinh ethnicity	0.86	0.35	1553
Rural areas	0.82	0.38	1553
Height-for-age z-score	-1.36	1.11	1553
Lao Cai in the fourth round	0.20	0.40	1553
Da Nang in the fourth round	0.17	0.38	1553
Hung Yen in the fourth round	0.20	0.40	1553
Phu Yen in the fourth round	0.22	0.41	1553
Ben Tre in the fourth round	0.20	0.40	1553
Other Province in the fourth round	0.01	0.07	1553
Mom's education: Up to Primary education	0.36	0.48	1553
Mom's education: Junior high school education	0.49	0.50	1553
Mom's education: High school education	0.09	0.29	1553
Mom's education: Post-secondary/vocational, University, Master's education	0.06	0.23	1553
Dad's education: Up to Primary education	0.31	0.46	1553
Dad's education: Junior high school education	0.46	0.50	1553
Dad's education: High school education	0.15	0.36	1553
Dad's education: Post-secondary/vocational, University, Master's education	0.07	0.26	1553
Mom's Job: Unemployed	0.16	0.37	1553

<sup>a</sup> Lao Cai residence (the poorest of the five provinces) is baseline for the regional dummy variables in the models

<sup>b</sup> “Other province” indicates provinces not included in the original Young Lives sampling, these children moved out of the five study provinces.

<sup>c</sup> Up to primary education (including literacy and religious education) is baseline for the mother and father’s education dummy variables in the models

Table 4: OLS Regression Estimates

	Model 1	Model 2
	Verbal Cognition (PPVT) <sup>a</sup>	Math
Years spent at preschool	0.60* (1.89)	0.77 (1.59)
Age of child in months	0.09 (1.17)	0.15 (1.38)
Sex of members	-0.76 (-1.57)	1.76** (2.38)
Wealth index	9.34*** (4.66)	19.19*** (6.28)
Kinh ethnicity	6.98*** (7.09)	7.25*** (4.82)
Rural areas	-1.88 (-0.70)	-5.76 (-1.40)
Height-for-age z-score	1.38*** (5.47)	1.19*** (3.10)
Mom's Job: Unemployed	0.23 (0.33)	-2.84*** (-2.68)
Observations	1553	1553
R-squared	0.216	0.264

*t* statistics in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

<sup>a</sup> Coefficients represent the estimated impact in percentage points of one year of preschool on test scores

<sup>b</sup> Models also include dummy variables for region and parental education levels

<sup>c</sup> PPVT is a vocabulary test widely used as an indicator of cognitive development

Table 5: 2SLS Regression Estimates

	Model 3 Verbal Cognition (PPVT)	Model 4 Math
Years spent at preschool	30.80*** (3.57)	21.75*** (2.91)
Age of child in months	-2.13*** (-3.23)	-1.39** (-2.43)
Sex of members	-1.34 (-1.05)	1.36 (1.23)
Wealth index	1.59 (0.28)	13.81*** (2.81)
Kinh ethnicity	-1.63 (-0.46)	1.26 (0.41)
Rural areas	2.01 (0.28)	-3.05 (-0.50)
Height-for-age z-score	0.13 (0.18)	0.33 (0.51)
Mom's Job: Unemployed	0.15 (0.08)	-2.89* (-1.85)
Observations	1553	1553

*t* statistics in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

<sup>a</sup> Coefficients represent the estimated impact in percentage points of one year of preschool on test scores

<sup>b</sup> Models also include dummy variables for region and parental education level

<sup>c</sup> Instruments: Time travelling to district capital.

Table 6: Diagnostic tests - 2SLS Model for Percentage of Correct answers in PPVT Test

<b>Endogeneity test</b>	
Durbin $\chi^2$ statistic	85.038
p-value for Durbin $\chi^2$ statistic	0.000
Wu–Hausman F statistic	88.748
p-value for Wu–Hausman F statistic	0.000
Degrees of freedom	1
Denominator degrees of freedom for Wu–Hausman F	1532
<b>F statistics</b>	
Minimum eigenvalue statistic (F statistic)	14.246

Table 7: Diagnostic tests - 2SLS Model for Percentage of Correct answers in Math Test

<b>Endogeneity test</b>	
Durbin $\chi^2$ statistic	17.602
p-value for Durbin $\chi^2$ statistic	0.000
Wu–Hausman F statistic	17.563
p-value for Wu–Hausman F statistic	0.000
Degrees of freedom	1
Denominator degrees of freedom for Wu–Hausman F	1532
<b>F statistics</b>	
Minimum eigenvalue statistic (F statistic)	14.246

Table 8: First stage regression results – 2SLS models

	Preschool education
Distance to the capital of the district?	-0.005*** (-3.77)
Age of child in months	0.073*** (13.15)
Sex of members	0.024 (0.62)
Wealth index	0.139 (0.85)
Kinh ethnicity	0.247*** (3.13)
Rural areas	-0.129 (-0.60)
Height-for-age z-score	0.038* (1.87)
Mom's Job: Unemployed	-0.011 (-0.19)
Observations	1553
R-squared	0.384

*t* statistics in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

<sup>a</sup> Models also include dummy variables for region and parental education level

Table 9: 2SLS Regression Estimates in Standard Deviations

	Verbal Cognition (PPVT)	Math
Years spent at preschool (standardized)	2.78*** (3.57)	1.25*** (2.91)
Age of child in months (standardized)	-0.73*** (-3.23)	-0.30** (-2.43)
Sex of members (standardized)	-0.06 (-1.05)	0.04 (1.23)
Wealth index (standardized)	0.03 (0.28)	0.14*** (2.81)
Kinh ethnicity (standardized)	-0.05 (-0.46)	0.03 (0.41)
Rural areas (standardized)	0.07 (0.28)	-0.07 (-0.50)
Height-for-age z-score (standardized)	0.01 (0.18)	0.02 (0.51)
Mom's Job: Unemployed (standardized)	0.01 (0.08)	-0.06* (-1.85)
Observations	1553	1553

Coefficients in Standard Deviations; *t* statistics in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

<sup>a</sup> Models also include dummy variables for region and parental education level