

Contrasting Experiences

Understanding the Longer-Term Impact of Improving Access to Preschool Education in Rural Indonesia

Amer Hasan

Haeil Jung

Angela Kinnell

Amelia Maika

Nozomi Nakajima

Menno Pradhan



WORLD BANK GROUP

Education Global Practice

November 2019

Abstract

This paper examines the longer-term impact of a project that expanded access to playgroup services in rural Indonesia. It compares the outcomes of two cohorts of children who were exposed to the same intervention at different points in time. One cohort was eligible to access playgroups during the first year of a five-year project cycle, beginning at age four. The other cohort became eligible to access these services during the third year, beginning at age three. The younger cohort was more likely to be exposed to playgroups for longer and at age-appropriate times relative to the older cohort. The paper finds that enrollment rates and

enrollment duration in preprimary education increased for both cohorts, but the enrollment effects were larger for the younger cohort. In terms of child development outcomes, there were short term effects at age five that did not last until age eight, for both cohorts. The data reveal that the younger cohort had substantially higher test scores during the early grades of primary school, relative to the older cohort. To unpack why the two cohorts experienced different longer-term outcomes, the paper provides evidence of changes that transpired in the operating conditions of the playgroups over time.

This paper is a product of the Education Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at ahasan1@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Contrasting Experiences: Understanding the Longer-Term Impact of Improving Access to Preschool Education in Rural Indonesia

Authors*

Amer Hasan (World Bank)

Haeil Jung (Korea University)

Angela Kinnell (University of Adelaide)

Amelia Maika (Gadjah Mada University)

Nozomi Nakajima (Harvard University)

Menno Pradhan (University of Amsterdam and VU University)

JEL Codes

I24, I25

Keywords

Early Childhood Education, Child Development, Treatment impacts, Differential effects

*Authors are listed alphabetically.

Acknowledgments

We would like to thank Dedy Junaedi, Upik Sabainingrum, Anas Sutisna, Lulus Kusbudiharjo and Mulyana for managing the fieldwork. We are grateful to Sophie Naudeau, Alaka Holla and Halsey Rogers for comments on an earlier version. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Funding

This work was supported by the Strategic Impact Evaluation Fund (TF0A0234) at the World Bank. Haeil Jung acknowledges that this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A3A2924956).

Declarations of interest

None.

1. Introduction

A growing body of research shows that a child's early life has consequences for later life outcomes in education (Bhutta et al., 2002; Brennan et al., 2012; Duncan et al., 2007; Feinstein and Duckworth, 2006; Melhuish et al., 2008; Moser et al., 2012), health (Hertzman, 2013), and social capital (Moffitt et al., 2011). Healthy child development is an enabler of human capability allowing children to reach physical maturity and participate productively in economic, social and civic life (Conti and Heckman, 2012; Sen, 1999). Many of the problems arising in early childhood have social and financial costs that cumulatively represent a considerable drain on a country's resources (Feinstein and Duckworth, 2006; Victora et al., 2008).

High-quality preschool programs provide an opportunity to mitigate the risk factors that many young children face (Barnett, 2011; Duncan et al., 2007; Duncan and Magnuson, 2013; Heckman, 2006; Shonkoff and Phillips, 2000). Given the high-risk factors faced by children growing up in middle- and low-income countries (Engle et al., 2011), the effectiveness of preschool programs is likely to be large for these children. However, much of the evidence base on the long-term impacts of preschools has focused on three "iconic" projects in the United States: Perry Preschool, Abecedarian, and the Nurse Family Partnerships (Shonkoff, 2014). These studies examined high-intensity interventions in the United States that were run with small sample sizes in the late 1960s to early 1970s. While all have had longitudinal follow-ups, few interventions have been implemented since that match either the fidelity or the intensity of these interventions.

In developing country settings, rigorous evaluations of preschool programs have emerged in the last decade. Results have ranged from no effect in some settings (Bouguen, Filmer, Macours and Naudeau, 2017) to positive effects in others (Brinkman et al., 2017; Jung and Hasan, 2016; Burger, 2010; Martinez, Naudeau and Pereira, 2012; Nores and Barnett, 2010). Systematic reviews of preschool interventions in developing countries reinforce the wide range of impacts seen in international settings (Tanner et al., 2015). The contrasting evidence from different settings has led some to question whether early childhood education can even have lasting impacts (Stevens and English, 2016).

In this paper, we contribute to the growing literature on the longer-term impacts of preschools in developing countries. We examine the longer-term impacts of early childhood education in a developing country setting using data on two cohorts of children from the Indonesia Early Childhood Education and Development (ECED) Project. The project lasted five years and expanded access to community-based preschools by providing block grants to villages to establish up to two playgroups (targeted for ages 3-4), providing teacher training, and raising community awareness about the importance of early education.¹ The *older cohort* in our study was age 4 and the *younger cohort* was age 1 when the project began.

The older and younger cohorts differed in terms of (1) the phase of project implementation when they were exposed to playgroups and (2) the length of exposure to playgroups at the appropriate age. On the first distinction, the older cohort was exposed to playgroups during the project's first year since they were already age 4 at the start of the project. In contrast, the younger cohort was exposed to playgroups during the project's third year when they turned 3 years old. As a result, the two cohorts experienced the same project at different phases of implementation. On the second distinction, the older cohort was at the upper-end of the appropriate age for playgroup attendance (age 3-4) when they were exposed to the project. In contrast, the younger cohort was only 1 year old when the project started, which meant that

¹ In practice, at the start of the project, the playgroups allowed enrollment for children ages 3-6. See Brinkman et al, 2017a for further discussion.

they had an opportunity to enroll in playgroups at the appropriate age, starting at age 3 and continuing through age 4. Thus, the two cohorts differed in their likelihood of being exposed to playgroups at the appropriate age.

Our data allow us to track the development outcomes for both cohorts at ages 5 and 8. To control for children's baseline development, we use measures of child development before exposure to playgroups. For the younger cohort, we have data on child development measures from when they were 1 and 2 years old. For the older cohort, we have data on child development measures from when they were 4 years old. We use a comprehensive set of child development outcomes that measure both cognitive and socio-emotional development. In addition, we capture children's performance on a test of language, mathematics and abstract reasoning in primary school. Together, these measures allow us to trace out early development on a variety of dimensions.

We build upon previous work by Brinkman et al. (2017b), which analyzed the impact of the project on the older cohort. In this paper, we conduct three new analyses. First, we present child development outcomes for the younger cohort at ages 5 and 8. Second, we present new results for the older cohort using primary school test scores. Third, we contrast the developmental trajectories between the two cohorts. We explore how features of project implementation as well as different rates of exposure to playgroups at the appropriate age may have contributed to the contrasting results between the two cohorts.

This paper also contributes to research on the generalizability of impact evaluations in early childhood education. Preschool interventions in developing countries have produced a wide range of effect sizes, prompting both researchers and policy makers to question the generalizability of these findings (Tanner et al., 2015). In particular, it is difficult to understand what causes these variations since each preschool program is different in terms of what the program provides, how the program is implemented and where the program is carried out. In this paper, we are able to hold constant the program content (what) and the study context (where) but vary the implementation (how) experienced by two cohorts of children. In doing so, we document the extent to which program impacts may vary as a result of differences in project implementation.

This paper proceeds as follows: The second section describes the Indonesia Early Childhood Education and Development (ECED) Project. The third section explains the data we use in our analysis. The empirical strategy is described in the fourth section. The fifth section provides the empirical results and the sixth section compares the impacts on the younger cohort and older cohorts. The seventh section presents a cost-benefit analysis based on our intent-to-treat estimates of additional years of schooling completed by those living in project villages. The final section concludes with a discussion of the findings and their implications for future work.

2. Indonesia Early Childhood Education and Development (ECED) Project

The goals of the Indonesia ECED Project were to increase access to early childhood services and to increase children's school readiness in rural villages. The project consisted of three components. First, a community facilitator raised awareness about the importance of early childhood services and shared information about how to prepare a proposal for the block grants available through the project.

Second, block grants were provided to each village, in the amount of USD 18,000 per village over three years. Villages could use the grant to establish or support up to two early childhood education centers. No more than 20 percent of the grant could be spent on new infrastructure. The most common form of services established were playgroups, which are early childhood education services catered to children ages 3 and 4 before they enroll in

kindergarten at ages 5 and 6. Playgroups are typically half-day programs that meet every other day while kindergartens typically meet daily.

Third, the project included a component that provided 200 hours of teacher training for up to two teachers per project playgroup. Teachers were predominantly mothers in the villages, some of whom had prior work experience in health and education services.

Thus, the “treatment” evaluated in this impact evaluation refers to this package of interventions provided by the ECED project – a community facilitator, block grants to establish playgroups, and teacher training (Hasan, Hyson and Chang, 2013).

The project was implemented in relatively poor villages in rural Indonesia. Of the 442 districts in the country, 50 poor districts were selected based on having high poverty rates, low enrollment rates in early childhood education, and low Human Development Index rankings. Within each district, 60 priority villages were identified based on their poverty rate, population size, and willingness to participate in the project. Overall, the Indonesia ECED Project was implemented in 3,000 villages.²

According to project documents, the cost per child for the project’s package of interventions was about US\$30 per year. This estimate excludes any voluntary contributions from the villages to the project. Villages often made available the land on which playgroups were housed. In contrast, other early childhood programs range in cost from US\$37 per child in India to US\$52 per child in Mexico to US\$66 per child in Brazil—suggesting that this package was slightly less costly.³

Evaluation design

The impact evaluation of the project focuses on 310 villages in 9 districts that were part of the project. There are 218 treatment villages and 92 comparison villages. The treatment villages were randomly assigned to two batches. 105 villages received the project first in 2009 (referred to as *batch 1*) and 113 villages received the project 11 months later in 2010 (referred to as *batch 2*). The comparison villages never received the project. The district governments selected the comparison villages on the basis of having similar poverty levels to the treatment villages.⁴ Comparison villages were therefore not randomly assigned.

In each treatment and comparison village, approximately 10 households with a 1 year-old child (who became the younger cohort) and approximately 10 households with a 4 year-old child (who became the older cohort) were randomly selected for evaluation.⁵ Thus, the impact evaluation follows these two separate cohorts of children who were able to access the playgroups provided by the project at different time points, based on their age, and when the project was at different stages of maturity. The timeline in Figure 1 below shows the timing of the project and the ages of the two cohorts.

An earlier paper by Brinkman et al. (2017b) documented the impacts of the project on the older cohort which, based on their age, was eligible to enroll in playgroups established under the project as soon as the project was implemented in 2009 and 2010. It employed instrumental variables and difference-in-differences models to determine the impacts on these children when they were aged 5 and 8 respectively. The paper found that while the intervention raised enrollment rates and durations of enrollment, there was little impact on child development. The two models corresponded to different durations of project exposure. The

² For further details please see Pradhan et al. (2013).

³ See for instance the estimates quoted in Barnett (1997) and Evans, Myers and Ilfeld (2000).

⁴ Appendix 7-9 in Hasan, Hyson and Chang (2013) document that these villages are well balanced on a range of observable characteristics.

⁵ Only 32 children from the two cohorts are siblings to each other.

difference-in-differences model captured greater exposure and showed that there were modest and sustained impacts on child development, especially for children from more disadvantaged backgrounds.

The present paper focuses on the younger cohort who, based on their age, became eligible to enroll in playgroups later in the project's implementation, in 2011 and 2012. Similar to the previous analysis for the older cohort (Brinkman et al. 2017b), this paper estimates the impact of the intervention for the younger cohort at age 5 and at age 8 in terms of enrollment in pre-primary education and child development outcomes. We also capitalize on the two cohorts that experienced the project to contrast the impact estimates for the younger cohort with those of the older cohort. For the older cohort, in addition to estimates presented in Brinkman et al. (2017b), we report new estimates of the impact of exposure to early childhood education and development services on test scores in language (Bahasa Indonesia) conducted when the children were aged 8 and in the early grades of primary school.

Figure 1: Age of cohort, eligibility for various early childhood education, and project implementation phase for each survey round

Year	2009	2010	2011	2012	2013	2014	2015	2016
Survey	Round 1	Round 2	Round 3			Round 4		
Project phase	Start of project				End of project			
1-year-old cohort	1	2	3	4	5	6	7	8
Type of service child is eligible to enroll in	Children are too young to enroll in playgroups		Children are eligible to enroll in playgroups		Children are eligible to enroll in kindergartens		Children are eligible to enroll in primary school	
4-year-old cohort	4	5	6	7	8	9	10	11
Type of service child is eligible to enroll in	Children are eligible to enroll in playgroups	Children are eligible to enroll in kindergartens		Children are eligible to enroll in primary school		<i>These children were no longer surveyed but would have been in primary school</i>		

Note: Figure depicts ages of the two cohorts studied and what types of services they are eligible for at each age. Note: Bold numbers denote that the children in that cohort were surveyed at that age/year.

Source: Authors

3. Data

The main analyses in this paper use data on the younger cohort collected in 2013 (at age 5) and in 2016 (at age 8). The key outcomes of interest are (i) enrollment in different types of early childhood education services; (ii) child development outcomes using the Early Development Instrument (EDI); and (iii) tests scores in early grades of primary school.

We measure enrollment in three types of early childhood education services. The first are project playgroups, those established under the project's block grant. The second are non-project playgroups, which refer to all other playgroup services that exist in the communities. The third are kindergartens, which are early childhood education programs catering to children before they enroll in primary school. We collected information about enrollment in each type of service by collecting a retrospective enrollment history for each academic year from 2008 to the survey year from the primary caregiver of the child.

The EDI measures children's school readiness across five domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge (Janus and Offord, 2007). The EDI has been validated and tested for reliability. Overall, the construct validity, predictive validity, and inter-rater reliability of the EDI in Indonesia are comparable to that found in other countries, making the EDI a suitable instrument for measuring school readiness in Indonesia (Brinkman et al. 2016). In this paper, we use the short-form of the caregiver-rated EDI.⁶ We standardized the variables for each EDI domain to have a mean of 0 and standard deviation of 1, using the mean and standard deviation of the comparison group in the younger cohort.

A school-based test was developed for this evaluation based on learning standards in Indonesian schools. Children were assessed in a classroom under the guidance of a member of the data collection team and no classroom teachers were present. The tests were divided into three parts: language (Bahasa Indonesia), mathematics, and abstract reasoning. The language test consisted of two sections. The first section (*match pictures*) evaluated children's phonological awareness (i.e., whether they can match pictures that start with a given sound) and letter recognition (i.e., whether they can match pictures that start with a given letter). The second section (*mention objects*) assessed children's vocabulary skills (i.e., whether they can name the word associated with a given image). The mathematics test included two sections. The first section (*summation*) evaluated children's ability to add and subtract (i.e., whether they can add to or subtract away from a set of objects). The second section (*order numbers*) assessed children's ability to recognize patterns (i.e., whether they can order one- to two- digit numbers in ascending and descending order). The abstract reasoning section was modeled on the Raven's Progressive Matrices. Students were presented with an image that was missing a small section and asked to select the missing pieces from six options, based on color, pattern, and orientation. There were two versions of the overall test; a shorter test for 6 and 7 year-olds and a longer test for 8 and 9 year-olds. In this paper, we use the common set of items that were included in both versions of the test. We standardized the variables for each test domain to have a mean of 0 and standard deviation of 1, using the mean and standard deviation of the comparison group in the younger cohort.

Baseline measures of child development were collected for the younger cohort when they were aged 1 and 2. The EDI is not an appropriate test for that age group so instead a measure of child development developed by the University of San Carlos Office of Population

⁶ We use the short-form to match previously published estimates in Brinkman et al 2017b with which we compare this paper's findings. Results using the long-form of the EDI are qualitatively similar and are available upon request.

Studies that measured skills similar to those in the EDI was used. These measures were collected by asking the child’s primary caregiver whether the child is usually able to demonstrate various skills. Specifically, we directly observed (or, with younger or reluctant children, asked the mother about) children’s gross and fine motor skills, language, cognitive and socio-emotional development. In one set of questions, children were asked to demonstrate their ability to perform a specified skill. When the child did not want to demonstrate this skill, the mother was asked if the child was usually able to do it. In another set of questions, the mother was asked directly whether their child could perform a particular activity. For these skills, the child was never asked to do a demonstration. In all cases, higher values indicate better developmental outcomes (Office of Population Studies, 2005). We standardized each of the variables to have a mean of 0 and standard deviation of 1, using the mean and standard deviation of the comparison group in the younger cohort.

Although all children in the sample are from poor, rural areas, we measured the relative wealth of children’s households. Households were asked if they owned any of the following items: radio, television, refrigerator, bicycle, motorcycle, car, mobile phone, and livestock. They were also asked about the materials used to construct the floor, walls and roof of their homes. Households were also asked if they had access to electricity in their homes and whether they received government assistance. Using principal components analysis on these items, we constructed a single index of household wealth. We standardized the variable to have a mean of 0 and standard deviation of 1, using the younger cohort’s mean and standard deviation.⁷

As a measure of the child-parent relationship, we collected self-reports from the primary caregiver on how often they used various parenting practices related to their warmth, consistency, and hostility. The questionnaire was adapted from the Longitudinal Study of Australian Children (Zubrick et al. 2008). Higher total parenting scores correspond to higher levels of warmth and consistency, and lower levels of hostility. We standardized the variable to have a mean of 0 and standard deviation of 1, using the younger cohort’s mean and standard deviation.

4. Empirical Strategy

To estimate the causal effect of the project on the younger cohort, we would ideally compare the change in outcomes between age 1 and age 5 (or between age 1 and age 8) for children in the treatment villages, relative to the change in outcomes for children in the comparison villages (i.e., a difference-in-differences approach). However, children in the younger cohort were too young to have baseline measures of enrollment in early childhood education as they were not yet age eligible. These children were also too young to have baseline measures of the EDI as they were not old enough for the instrument.

Instead, we evaluate the impact of the project using the following regression specification:

$$y_{ijt=2013} = \alpha_0 + \alpha_1 T_j + X_{ijt=2013} \beta + V_{ijt=2009} \delta + \varepsilon_{ijt} \quad (1)$$

$$y_{ijt=2016} = \alpha_0 + \alpha_1 T_j + X_{ijt=2016} \beta + V_{ijt=2009} \delta + \varepsilon_{ijt} \quad (2)$$

⁷ A comparison of assets ownership by households in the evaluation sample with that of the rural sub-sample of the SUSENAS (a nationally representative household survey) suggests average rates of asset ownership and education levels are by and large similar between the two. See also Hasan, Hyson and Chang (eds.) 2013.

where y_{ijt} is the outcome measure of child i in village j at time t . T_j is an indicator for whether the village is treatment or comparison, X_{ijt} are time varying covariates (child's age, household size, household wealth index, and parenting score) and $V_{ijt=2009}$ are time invariant covariates (child's gender, whether the child's mother completed primary education or less, and baseline measures of child development). The key coefficient of interest is the treatment effect, α_1 . Equation (1) is the specification for 2013, which examines the effect of the intervention on enrollment rates and EDI at age 5. Equation (2) estimates the impact in 2016, which examines the effect of the project on enrollment rates, EDI, and test scores at age 8.

Our key identifying assumption is that the time varying and invariant covariates in our regression model fully account for any differences between children in the treatment villages and children in the comparison villages that are not due to treatment assignment.

We also examine the heterogeneity of the treatment effect across household wealth and parenting practices. We re-run our regression model separately for children with baseline household wealth below the sample mean (*poor*) and for children with baseline household wealth above the sample mean (*non-poor*). Similarly, we re-run our regression model separately for children with baseline parenting scores below (*low parenting score*) and above (*high parenting score*) the sample mean.

In Table 1, for the younger cohort, we show the summary statistics of child and family characteristics in treatment and comparison villages at baseline (2009). Columns 1 and 2 separately present the means and standard deviation for villages that received the project early (batch 1) and those that received the project later (batch 2), and column 3 presents these statistics for a sample that combines all treated villages together. Column 4 reports these statistics for the comparison villages that never received the project but were chosen because of their similarity to treatment villages. Column 5 reports the differences between villages that received the project early or late while column 6 reports the differences between treatment and comparison villages. In both of these cases the estimates reported are the results of a regression with standard errors clustered at the village level.

Column 3 shows that at baseline, younger cohort children in the treatment villages were around 1.5 years old. On average, children lived in households with wealth z-scores and parenting z-scores slightly below the sample mean. About half of the cohort's mothers had primary education or less and about half of the children were male. The mean body mass index (BMI) of the children in the younger cohort was 14.6 kg/m² and on a range of cognitive, fine motor, gross motor and language skills their scores were slightly below the sample mean.

Column 5 reports that there are no differences in these child and family characteristics between the two batches of treated villages. As a result, we examine batch 1 and batch 2 villages collectively as treatment villages in our regression specification in equations (1) and (2).

Column 6 shows that at baseline, treatment and comparison villages were generally similar in terms of various child and family characteristics. However, three variables showed statistically significant differences. On average, children in treatment villages lived in households with 0.259 fewer people than children in comparison villages. While this mean difference is statistically significant, the magnitude is small. We also find that children in treatment villages scored lower in measures of baseline cognitive skills (-0.143 S.D.) and gross motor skills (-0.153 S.D.). Thus, we control for these baseline differences in child development in our estimation of the treatment effect.

Table 1. Summary statistics for the younger cohort at baseline (2009)

	Treatment			Comparison	Differences	
	Early (Batch 1)	Late (Batch 2)	Both (Batch 1 & 2)		Early – Late	Both – Comparison
	(1)	(2)	(3)		(5)	(6)
Age (years)	1.520 (0.287)	1.499 (0.286)	1.509 (0.286)	1.508 (0.286)	-0.021 (0.012)	0.001 (0.013)
Household size	4.678 (1.529)	4.729 (1.568)	4.704 (1.549)	4.964 (1.705)	0.052 (0.095)	-0.259* (0.102)
Wealth z-score (S.D.)	-0.0125 (1.036)	-0.0301 (0.962)	-0.0217 (0.998)	0.0520 (1.003)	-0.018 (0.082)	-0.074 (0.073)
Parenting z-score (S.D.)	-0.0120 (1.006)	-0.0316 (0.968)	-0.0222 (0.986)	0.0532 (1.031)	-0.020 (0.070)	-0.075 (0.073)
Mother's edu is primary or less (1 = Yes)	0.512 (0.500)	0.514 (0.500)	0.513 (0.500)	0.504 (0.500)	0.002 (0.029)	0.008 (0.030)
Male (1 = Yes)	0.491 (0.500)	0.516 (0.500)	0.504 (0.500)	0.525 (0.500)	0.025 (0.023)	-0.021 (0.019)
Body Mass Index (BMI)	14.62 (2.078)	14.53 (2.117)	14.57 (2.098)	14.66 (2.097)	-0.092 (0.107)	-0.091 (0.102)
Cognitive skills (S.D.)	-0.0979 (1.055)	0.00907 (0.982)	-0.0421 (1.019)	0.101 (0.947)	0.107 (0.076)	-0.143* (0.061)
Fine motor skills (S.D.)	-0.0323 (0.988)	-0.00911 (1.045)	-0.0202 (1.018)	0.0483 (0.954)	0.023 (0.059)	-0.069 (0.055)
Gross motor skills (S.D.)	-0.0798 (1.013)	-0.0134 (1.017)	-0.0452 (1.015)	0.108 (0.954)	0.066 (0.057)	-0.153** (0.047)
Language skills (S.D.)	-0.0216 (0.984)	-0.00990 (1.031)	-0.0155 (1.008)	0.0371 (0.979)	0.012 (0.058)	-0.053 (0.052)
Observations	1042	1137	2179	910		

p<0.001***; p<0.01**; p<0.05*

Note: Early (Batch 1) villages received the project first in 2009 and late (batch 2) villages received the project later in 2010. Comparison villages never received the project. Standard deviation in parentheses in columns (1) to (4). Standard errors clustered at village level in parentheses in columns (5) to (6).

5. Results

The intent-to-treat impact estimates for the younger cohort are presented in Tables 2 to 4.⁸ In each table, column 1 presents the estimates for all children in the cohort, columns 2 and 3 separately estimate the impacts by relative household wealth at baseline, and columns 4 and 5 separately estimate the impacts by relative parenting score at baseline.⁹ When interpreting these results, it is important to note that the counterfactual is children living in comparison villages, which may or may not have other preschool services (i.e., non-project playgroups and kindergartens) available in their villages.

Table 2 presents impacts on enrollment rates and duration. In 2013 (at age 5), children in treatment villages were 49.9 percentage points more likely to report ever being enrolled in project playgroups compared to children from comparison villages. The treatment effect on enrollment rate was similar in 2016 when the children were age 8. Moreover, the effects are largely consistent across sub-samples.¹⁰ As expected, there was virtually no enrollment in project playgroups reported by children in comparison villages. In contrast to the increase in enrollment in project playgroups, children from treatment villages were 22 percentage points less likely to enroll in non-project playgroups relative to a 33.4 percent enrollment rate among children in comparison group villages. These estimates were fairly similar at ages 5 and 8. Finally, there was no difference in enrollment in kindergarten between treatment and comparison villages by age 5. However, by age 8, children from treatment villages were 8.6 percentage points less likely to have ever enrolled in kindergartens compared to children from comparison villages.

The results for months of enrollment are largely consistent with our findings for enrollment rate. The project increased children's enrollment duration in project playgroups, decreased enrollment duration in non-project playgroups, and kept enrollment duration in kindergarten unaffected.

One way to interpret the months of enrollment in a project playgroup is that it is a measure of "take-up" of the project. The average take-up of the project playgroup in treatment villages was 7.781 months by age 5 and 8.683 months by age 8.¹¹ These large effects on months of enrollment in project playgroups hold across wealth and parenting sub-group analyses. Specifically, compared to poor children in comparison villages, poor children in project villages enrolled in 7.3 more months of playgroup. Non-poor children enrolled for 8.1 more months. However, there is no evidence to suggest that these point estimates are different from each other.¹² The project thus had equally large enrollment effects for children from poor and non-poor households. Similarly, we do not find treatment effect variation between children from households with high and low parenting scores.

On average, children in treatment villages were enrolled in non-project playgroups for 3 fewer months than their peers in comparison villages. At age 5, this decrease in enrollment duration in non-project playgroups was significantly more pronounced for children from non-poor households. However, by age 8, there was no treatment effect variation in non-project

⁸ Results by batch are reported in Appendix Tables 1-3.

⁹ In both cases – wealth and parenting practices – we split the sample into those above the mean and those below the mean.

¹⁰ Appendix Table 4 provides results which show that the differences between groups – either poor versus non-poor or low versus high parenting practices are not statistically different from each other.

¹¹ Since all of our impact estimates focus on intent-to-treat, these take-up figures do not affect the interpretation of our results. The small non-zero estimates of enrollment in project centers among children from non-project villages are possible in those few cases where households lived near enough to a treatment community and project playgroup. However, this was rare.

¹² See Appendix Table 4.

playgroup enrollment between poor and non-poor children.

Finally, we do not find significant differences in enrollment duration in kindergarten between treatment and comparison villages, either in 2013 or in 2016. On average, all children in the sample seem to have completed about 5 months of kindergarten by 2013 and about 10 months of kindergarten by 2016 – with no substantial variation by household wealth or parenting.

Table 2. Impact on enrollment outcomes for the younger cohort

Outcome	Survey Year		All	Poor	Non-poor	Low Parenting Score	High Parenting Score
			(1)	(2)	(3)	(4)	(5)
Ever enrolled in project playgroup until survey year	2013	Coeff.	0.499***	0.494***	0.504***	0.501***	0.498***
		(S.E.)	(0.022)	(0.029)	(0.024)	(0.025)	(0.026)
		Comp. mean	0.00732	0.0116	0.00421	0.00699	0.00767
	2016	Obs.	2,778	1,201	1,577	1,540	1,238
		Coeff.	0.498***	0.539***	0.470***	0.466***	0.532***
		(S.E.)	(0.023)	(0.029)	(0.027)	(0.027)	(0.026)
Ever enrolled in non-project playgroup until survey year	2013	Comp. mean	0.0673	0.0499	0.0802	0.0686	0.0659
		Obs.	2,894	1,289	1,605	1,834	1,060
		Coeff.	-0.220***	-0.193***	-0.241***	-0.215***	-0.223***
	2016	(S.E.)	(0.037)	(0.044)	(0.042)	(0.041)	(0.043)
		Comp. mean	0.334	0.284	0.371	0.324	0.345
		Obs.	2,778	1,201	1,577	1,540	1,238
Ever enrolled in kindergarten until survey year	2013	Coeff.	-0.221***	-0.190***	-0.241***	-0.209***	-0.232***
		(S.E.)	(0.037)	(0.045)	(0.042)	(0.041)	(0.045)
		Comp. mean	0.407	0.338	0.459	0.389	0.427
	2016	Obs.	2,894	1,289	1,605	1,834	1,060
		Coeff.	-0.075	-0.074	-0.085*	-0.090	-0.059
		(S.E.)	(0.040)	(0.053)	(0.041)	(0.048)	(0.043)
Months enrolled in project playgroup until survey year	2013	Comp. mean	0.532	0.441	0.598	0.522	0.542
		Obs.	2,778	1,201	1,577	1,540	1,238
		Coeff.	-0.086*	-0.068	-0.098*	-0.098*	-0.075
	2016	(S.E.)	(0.040)	(0.054)	(0.038)	(0.044)	(0.046)
		Comp. mean	0.743	0.670	0.796	0.735	0.751
		Obs.	2,894	1,289	1,605	1,834	1,060
Months enrolled in non-project playgroup until survey year	2013	Coeff.	7.781***	7.319***	8.129***	7.748***	7.843***
		(S.E.)	(0.396)	(0.511)	(0.447)	(0.454)	(0.469)
		Comp. mean	0.0622	0.122	0.0189	0.0559	0.0691
	2016	Obs.	2,778	1,201	1,577	1,540	1,238
		Coeff.	8.683***	8.840***	8.599***	8.365***	9.022***
		(S.E.)	(0.413)	(0.544)	(0.469)	(0.496)	(0.491)
Months enrolled in non-project playgroup until survey year	2013	Comp. mean	0.327	0.321	0.331	0.320	0.334
		Obs.	2,894	1,289	1,605	1,834	1,060
		Coeff.	-3.001***	-2.094***	-3.743***	-2.721***	-3.284***
	2016	(S.E.)	(0.637)	(0.571)	(0.844)	(0.629)	(0.798)
		Comp. mean	4.635	3.290	5.613	4.228	5.082
		Obs.	2,778	1,201	1,577	1,540	1,238
2016	Coeff.	-3.187***	-2.476***	-3.706***	-2.913***	-3.443***	
	(S.E.)	-0.64	(0.651)	(0.814)	(0.660)	(0.796)	
	Comp. mean	5.346	4.241	6.167	4.975	5.741	
Obs.	2,894	1,289	1,605	1,834	1,060		

Table 2. Impact on enrollment outcomes for the younger cohort

Outcome	Survey Year		All	Poor	Non-poor	Low Parenting Score	High Parenting Score
			(1)	(2)	(3)	(4)	(5)
Months enrolled in kindergarten until survey year	2013	Coeff.	-0.199	0.068	-0.516	-0.151	-0.273
		(S.E.)	(0.436)	(0.505)	(0.496)	(0.479)	(0.513)
		Comp. mean	4.983	3.777	5.859	4.555	5.453
	2016	Obs.	2,778	1,201	1,577	1,540	1,238
		Coeff.	-1.012	-0.589	-1.326	-1.365	-0.692
		(S.E.)	(0.663)	(0.844)	(0.683)	(0.722)	(0.767)
	Comp. mean	10.24	9.008	11.16	10.06	10.43	
	Obs.	2,894	1,289	1,605	1,834	1,060	

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at the village level are in parentheses.

“Comp. mean” refers to the comparison group mean for the outcome variable. Column (1) regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother’s education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language). Columns (2) and (3) regressions use the same controls as column (1) except they exclude household wealth. Columns (4) and (5) regressions use the same controls as column (1) except they exclude parenting practices.

Next, we turn to results on child development outcomes. Table 3 presents the impact estimates on the EDI at age 5 (2013) and age 8 (2016). Overall, we find a few positive impacts of the project on children’s developmental outcomes at age 5 but no positive impacts at age 8. At age 5, we estimate a 0.208 S.D. increase in scores on the physical health and well-being domain and a 0.115 S.D. increase in scores on the emotional maturity domain for the overall sample. However, these effects do not persist to age 8. We generally observe null effects in 2016, with one negative impact on the communication skills and general knowledge domain (-0.136 S.D). These results are largely consistent with the fade-out literature in early childhood education, where interventions aimed to reduce early childhood inequalities have large initial impacts but these dissipate over time (see for instance the studies included in Tanner et al., 2015). Our results suggest that the impact on emotional maturity may be concentrated among those classified as poor in our data.¹³ There are no other statistical differences across subgroups – either by wealth or by parenting practices.

One note of caution in interpreting our results is warranted. The data suggest that measurement error may contribute to the null effects on the EDI that we observe in 2016. In the raw histograms of the EDI domains presented in Appendix Figure 1, we find evidence of ceiling effects in 2016 that are not present in 2013. The exception in 2013 was ceiling effects on the communication skills and general knowledge domain, which may also explain the negative impact on this domain identified in 2016.

¹³ See Appendix Table 5.

Table 3. Impact on EDI outcomes for the younger cohort

Outcome	Survey Year		All	Poor	Non-poor	Low Parenting Score	High Parenting Score
Physical health and well-being (SD)	2013	Coeff.	0.208***	0.115	0.267***	0.151*	0.275***
		(S.E.)	(0.051)	(0.082)	(0.059)	(0.064)	(0.072)
		Comp. mean	-0.149	-0.118	-0.172	-0.191	-0.102
	Obs.	2,770	1,194	1,576	1,533	1,237	
	2016	Coeff.	0.023	0.113	-0.047	0.013	0.025
		(S.E.)	(0.065)	(0.103)	(0.060)	(0.085)	(0.069)
Comp. mean		0.00610	-0.170	0.136	-0.153	0.174	
Obs.	2,877	1,279	1,598	1,823	1,054		
Social competence (SD)	2013	Coeff.	0.018	0.080	-0.031	0.014	0.016
		(S.E.)	(0.053)	(0.082)	(0.058)	(0.056)	(0.079)
		Comp. mean	0.00112	-0.192	0.140	-0.139	0.157
	Obs.	2,769	1,192	1,577	1,534	1,235	
	2016	Coeff.	0.008	0.052	-0.023	-0.036	0.040
		(S.E.)	(0.049)	(0.067)	(0.058)	(0.070)	(0.059)
Comp. mean		0.0284	-0.0539	0.0892	-0.249	0.321	
Obs.	2,877	1,279	1,598	1,823	1,054		
Emotional maturity (SD)	2013	Coeff.	0.115*	0.241**	0.025	0.115	0.114
		(S.E.)	(0.056)	(0.082)	(0.057)	(0.076)	(0.059)
		Comp. mean	-0.110	-0.228	-0.0251	-0.333	0.138
	Obs.	2,770	1,193	1,577	1,534	1,236	
	2016	Coeff.	0.020	0.098	-0.042	-0.021	0.049
		(S.E.)	(0.046)	(0.061)	(0.058)	(0.064)	(0.059)
Comp. mean		-0.00663	-0.145	0.0955	-0.305	0.307	
Obs.	2,877	1,279	1,598	1,823	1,054		
Language and cognitive development (SD)	2013	Coeff.	0.073	0.098	0.043	0.098	0.051
		(S.E.)	(0.056)	(0.084)	(0.061)	(0.064)	(0.074)
		Comp. mean	-0.0502	-0.285	0.117	-0.154	0.0648
	Obs.	2,770	1,193	1,577	1,534	1,236	
	2016	Coeff.	0.080	0.078	0.090	0.085	0.067
		(S.E.)	(0.049)	(0.074)	(0.056)	(0.067)	(0.067)
Comp. mean		-0.0371	-0.110	0.0169	-0.103	0.0324	
Obs.	2,877	1,279	1,598	1,823	1,054		
Communication skills and general knowledge (SD)	2013	Coeff.	-0.005	-0.095	0.055	-0.069	0.067
		(S.E.)	(0.074)	(0.101)	(0.076)	(0.081)	(0.091)
		Comp. mean	0.0159	-0.0462	0.0603	0.00577	0.0271
	Obs.	2,771	1,194	1,577	1,534	1,237	
	2016	Coeff.	-0.136*	-0.145*	-0.123	-0.153	-0.127
		(S.E.)	(0.061)	(0.072)	(0.074)	(0.078)	(0.067)
Comp. mean		0.131	0.137	0.127	-0.0184	0.289	
Obs.	2,877	1,279	1,598	1,823	1,054		

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at the village level are in parentheses. "Comp. mean" refers to the comparison group mean for the outcome variable. Column (1) regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language). Columns (2) and (3) regressions use the same controls as column (1) except they exclude household wealth. Columns (4) and (5) regressions use the same controls as column (1) except they exclude parenting practices.

In Table 4, we find mixed results of the project on primary school test scores, which were measured at age 8 (2016). We find large positive effects (0.134 S.D.) on language items involved with selecting a picture whose name began with a different letter to other pictures

(*match picture*) but null effects on language tasks associated with writing the name of everyday items (*mention objects*). For the mathematics section, we similarly find large positive effects (0.125 S.D.) on tasks associated with ordering sequences of numbers from largest to smallest and vice versa (*order numbers*) but null effects on solving addition problems (*summation*). Finally, we find no impact on abstract reasoning. Overall, we do not detect treatment effect variation in the test score results, either by wealth or by parenting practices.¹⁴ As was the case for the EDI outcomes, test scores also seem to have ceiling effects associated with them, which may undermine our ability to detect certain effects.¹⁵

Table 4. Impact on primary school test scores for the younger cohort

Outcome		All	Poor	Non-poor	Low Parenting Score	High Parenting Score
Language - match picture (SD)	Coeff.	0.134*	0.029	0.213**	0.179**	0.080
	(S.E.)	(0.058)	(0.081)	(0.068)	(0.067)	(0.077)
	Comp. mean	-0.0827	-0.146	-0.0355	-0.163	0.00282
	Obs.	2,862	1,274	1,588	1,814	1,048
Language - mention objects (SD)	Coeff.	-0.027	-0.065	0.006	-0.030	-0.028
	(S.E.)	(0.053)	(0.067)	(0.065)	(0.062)	(0.063)
	Comp. mean	0.0333	-0.108	0.138	-0.0354	0.106
	Obs.	2,862	1,274	1,588	1,814	1,048
Math - summation (SD)	Coeff.	0.068	0.059	0.075	0.055	0.082
	(S.E.)	(0.054)	(0.077)	(0.061)	(0.067)	(0.068)
	Comp. mean	-0.0407	-0.166	0.0521	-0.0730	-0.00628
	Obs.	2,862	1,274	1,588	1,814	1,048
Math - order numbers (SD)	Coeff.	0.125*	0.108	0.143*	0.067	0.182*
	(S.E.)	(0.058)	(0.083)	(0.067)	(0.068)	(0.073)
	Comp. mean	-0.0828	-0.214	0.0143	-0.0905	-0.0745
	Obs.	2,862	1,274	1,588	1,814	1,048
Abstract reasoning (SD)	Coeff.	-0.022	-0.012	-0.032	-0.020	-0.023
	(S.E.)	(0.044)	(0.064)	(0.056)	(0.058)	(0.059)
	Comp. mean	0.0310	-0.102	0.130	-0.0317	0.0977
	Obs.	2,862	1,274	1,588	1,814	1,048

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at the village level are in parentheses. "Comp. mean" refers to the comparison group mean for the outcome variable. Column (1) regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language). Columns (2) and (3) regressions use the same controls as column (1) except they exclude household wealth. Columns (4) and (5) regressions use the same controls as column (1) except they exclude parenting practices.

¹⁴ Appendix Table 6.

¹⁵ See Appendix Figure 2.

6. Contrasting experiences – comparing project impacts for the younger and older cohorts

As described at the outset, the impact evaluation of the Indonesia ECED Project followed two cohorts of children. The focus of the paper so far has been on the younger cohort. The impacts of the intervention on the older cohort have previously been published (Brinkman et al. 2017b).

As shown in Figure 1, the data collection schedule meant that both cohorts of children were surveyed when they were age 5 and age 8. In this section, we contrast the impact estimates of these two cohorts side-by-side.¹⁶ Table 5 presents the impact estimates at age 5 and age 8 for each cohort. Columns 1 and 4 are the results for the younger cohort previously shown in Tables 2 to 4 in this paper. Columns 2 and 5 present the equivalent results for the older cohort, previously reported in Brinkman et al., 2017b. Column 6 presents test score results for the older cohort, which have not previously been published.¹⁷ Columns 3 and 7 show the results of a *t*-test comparing the differences in impact between the two cohorts.¹⁸

Overall, we find that the project impacts differed between the older and younger cohorts. The results are consistent with the fact that the *experiences* of the older and younger cohorts were different along two key dimensions: (1) the phase of implementation when they were exposed to project playgroups and (2) the length of exposure to project playgroups at the appropriate age (3-4 years old). On the first dimension, children in the older cohort were exposed to the project in its first year whereas children in the younger cohort were exposed to a more mature project in its third year of implementation. On the second dimension, children in the older cohort were capped (at most) to one year of age-appropriate enrollment in project playgroups because they were already age 4 when the project began. In contrast, children in the younger cohort were only age 1 when the project began so they were more likely to enroll in project playgroups at the appropriate-ages of 3 to 4.

Table 5 shows that the younger and older cohorts had very different rates of enrollment in various types of early childhood education. Children from the younger cohort in treatment villages were 49.9 percentage points more likely to enroll in project playgroups compared to children from comparison villages. In contrast, the estimate for the older cohort was only 15.3 percentage points. This corresponds to a difference of 34.6 percentage points when the two cohorts were age 5. By age 8, the difference in impacts between the two cohorts had declined somewhat but was still very large. The effect on project playgroup enrollment for children in the younger cohort was 22.9 percentage points higher than that for the children in the older cohort.

For enrollment in non-project playgroups, the impact was a 22 percentage point decline for the younger cohort compared to 4.4 percentage point decline for the older cohort. By age 5, children in the younger cohort were 17.6 percentage points less likely to be enrolled in non-project playgroups compared to the older cohort. These differences in estimates were fairly similar by age 8, with a 15.2 percentage point difference between the two cohorts.

For kindergarten enrollment, the point estimates for the two cohorts were not

¹⁶ See Appendix B for the empirical strategy used to estimate the impacts for the 4 year-old cohort.

¹⁷ These are obtained using the approach described in Appendix B. The items used to construct scores for the 4 year-old cohort are identical to those used for the 1 year-old cohort.

¹⁸ Our test statistic to examine whether the estimates from the two cohorts are statistically different is

$$\frac{\widehat{\beta}_{1yr} - \widehat{\beta}_{4yr}}{SE(\widehat{\beta}_{1yr} - \widehat{\beta}_{4yr})} = \frac{\widehat{\beta}_{1yr} - \widehat{\beta}_{4yr}}{\sqrt{Var(\widehat{\beta}_{1yr}) + Var(\widehat{\beta}_{4yr}) - 2Cov(\widehat{\beta}_{1yr}, \widehat{\beta}_{4yr})}} = \frac{\widehat{\beta}_{1yr} - \widehat{\beta}_{4yr}}{\sqrt{SE(\widehat{\beta}_{1yr})^2 + SE(\widehat{\beta}_{4yr})^2}} \text{ assuming}$$

$2Cov(\widehat{\beta}_{1yr}, \widehat{\beta}_{4yr}) = 0$. Since $Cov(\widehat{\beta}_{1yr}, \widehat{\beta}_{4yr})$ is typically positive, our assumption yields a conservative estimate of the test statistic.

distinguishable from each other – either by age 5 or by age 8. It is worth noting that by age 8, kindergarten enrollment had declined by 8.6 percentage points for the younger cohort and 13 percentage points for the older cohort. This suggests that the project may have encouraged children in treatment villages to substitute away from kindergarten relative to children in comparison villages, and this rate of substitution was not distinguishable across the two cohorts.

The younger cohort was also exposed to a higher dose of early childhood education than the older cohort. Relative to the older cohort, the younger cohort had accumulated 6 more months of project playgroups by age 5 and about 5 more months of project playgroups by age 8. This difference in enrollment duration between the two cohorts is likely driven by the fact that the older cohort was at the upper-end of the appropriate age for project playgroups at the beginning of the project. In contrast, the younger cohort was able to enroll in project playgroups at the appropriate age and remain in playgroups for longer.

The younger cohort enrolled in fewer months of non-project playgroups than the older cohort both by age 5 (3 months versus 0.6 months) and by age 8 (3.3 months compared to 0.9 months). Finally, we cannot reject the null hypothesis that the enrollment duration in kindergarten was significantly different across the two cohorts, both at age 5 and age 8.¹⁹

Looking across the three types of services, the younger cohort was more likely to be enrolled in project playgroups and more likely to be enrolled for longer than the older cohort. As we turn to the results on child development and test scores, this fact should be borne in mind.

At age 5, the impacts on EDI outcomes varied across the two cohorts. Specifically, the younger cohort experienced an improvement in physical health and well-being (0.208 standard deviations) and emotional maturity (0.115 standard deviations). In comparison, the older cohort showed improvements in the domain of social competence (0.223 standard deviations). In particular, the treatment effect was significantly larger for the younger cohort than the older cohort in terms of physical health and well-being. However, for social competence, the treatment effect was significantly larger for the older cohort than the younger cohort.

At age 8, the magnitudes of the treatment effect on the EDI domains were smaller for the younger cohort compared to the older cohort in every dimension. However, the treatment effects for the younger cohort were not statistically different from the treatment effects for the older cohort. This was true across all 5 domains of the EDI.²⁰

¹⁹ For the older cohort, the significant decline in kindergarten duration (-1.546) was consistent with the significant decline in kindergarten enrollment rate (-0.130). For the younger cohort, we did not find a significant decline in kindergarten duration but did find a significant decline in kindergarten enrollment rate (-0.086). This inconsistency for the younger cohort is due to the lack of precision in the estimates for months of enrollment. Overall, we interpret the results as evidence of substitution away from kindergarten in both cohorts.

²⁰ In the case of communication skills and general knowledge domain the data suggest that treated children are in fact doing worse than comparison children (0.16 standard deviations). One explanation for this result are the ceiling effects shown in Appendix Figure 1.

Table 5. Comparison of impact estimates for the two cohorts

	Treatment effect at age 5			Treatment effect at age 8			
	Younger cohort	Older cohort (JOLE results)	Difference. between cohorts	Younger cohort	Older cohort (JOLE results)	Older cohort (new results)	Difference. between cohorts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ever enrolled in project playgroup	0.499*** (0.022)	0.153*** (0.015)	0.346*** (0.027)	0.498*** (0.023)	0.269*** (0.018)		0.229*** (0.029)
Ever enrolled in non-project playgroup	-0.220*** (0.037)	-0.044** (0.014)	-0.176*** (0.040)	-0.221*** (0.037)	-0.069*** (0.017)		-0.152*** (0.041)
Ever enrolled in kindergarten	-0.075 (0.040)	-0.058* (0.026)	-0.017 (0.048)	-0.086* (0.040)	-0.130*** (0.028)		0.044 (0.049)
Months enrolled in project playgroup	7.781*** (0.396)	1.914*** (0.158)	5.867*** (0.426)	8.704*** (0.414)	3.855*** (0.254)		4.849*** (0.486)
Months enrolled in non-project playgroup	-3.001*** (0.637)	-0.647*** (0.156)	-2.354*** (0.656)	-3.274*** (0.645)	-0.943*** (0.238)		-2.331*** (0.688)
Months enrolled in kindergarten	-0.199 (0.436)	0.590* (0.270)	-0.789 (0.513)	-1.037 (0.664)	-1.546*** (0.444)		0.509 (0.799)
Physical health and well-being (SD)	0.208*** (0.051)	-0.026 (0.076)	0.234* (0.091)	0.002 (0.064)	0.104 (0.074)		-0.102 (0.098)
Social competence (SD)	0.018 (0.053)	0.223** (0.076)	-0.205* (0.093)	-0.012 (0.049)	0.024 (0.075)		-0.036 (0.09)
Emotional maturity (SD)	0.115* (0.056)	0.014 (0.071)	0.101 (0.090)	0.003 (0.045)	0.158* (0.068)		-0.155 (0.082)
Language and cognitive development (SD)	0.073 (0.056)	0.128 (0.070)	-0.055 (0.090)	0.035 (0.040)	0.056 (0.060)		-0.021 (0.072)
Communication and general knowledge (SD)	-0.005 (0.074)	0.075 (0.079)	-0.08 (0.108)	-0.160*** (0.061)	0.014 (0.132)		-0.174 (0.145)
Language - match picture (SD)				0.133** (0.058)		-0.100 (0.058)	0.233*** (0.082)
Language - mention objects (SD)				-0.027 (0.053)		-0.050 (0.054)	0.023 (0.076)

Table 5. Comparison of impact estimates for the two cohorts

	Treatment effect at age 5			Treatment effect at age 8			
	Younger cohort	Older cohort (JOLE results)	Difference. between cohorts	Younger cohort	Older cohort (JOLE results)	Older cohort (new results)	Difference. between cohorts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Math - summation (SD)				0.068 (0.054)		-0.148** (0.051)	0.216*** (0.074)
Math - order numbers (SD)				0.125** (0.058)		-0.155* (0.061)	0.280*** (0.084)
Abstract reasoning (SD)				-0.021 (0.044)		-0.124* (0.057)	0.103 (0.072)

p<0.001***; p<0.01**; p<0.05*

Note: Each cell is the result of a separate regression. Standard errors clustered at the village level are in parentheses. Regressions for the younger cohort control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language). The difference between the treatment effects for the younger and older cohorts are estimated using a Welch's t-test or unequal variance t-test.

Lastly, we compared how the cohorts fared in terms of tests of language, mathematics and abstract reasoning in the early grades of primary school. The children in the younger cohort who live in treatment villages scored 0.133 standard deviations higher than children in comparison villages on a language test that asks them to match pictures with words. They also score 0.125 standard deviations higher than children from comparison villages on a math test that asks them to order numbers. In contrast, for the older cohort, children exposed to the treatment have similar scores in language tests when compared to children from comparison villages but lower scores on summation (-0.148), number ordering (-0.155) and abstract reasoning (-0.124).

Overall, children in the younger cohort performed better across a range of dimensions tested – the picture matching (0.233 standard deviations), ordering numbers (0.28 standard deviations), and summation (0.216 standard deviations). These results suggest that children with greater exposure to project playgroups, that is, children in the younger cohort, performed better in tests administered in primary schools than those in the older cohort who had significantly less exposure to project playgroups.

7. Cost-benefit analysis

Before discussing these results in greater detail, we turn to another critical question: was the Indonesia ECED project a worthwhile investment? This section argues that it was. Comparable interventions in other countries range in cost from US\$37 per child in India to US\$289 in Colombia. The Indonesian project, on the other hand, cost approximately US\$27 per child (all amounts in 2014 dollars).

Using the actual number of children reached by the project (673,162 as at June 2013) and the actual observed increase in educational attainment (0.1 years on average for the older cohort and 0.7 years for the younger cohort) allows us to present a rudimentary cost-benefit analysis (Table 6). It uses a conservative set of estimates of rates of return to education: which range from 6.8-10.6 percent as estimated by Duflo (2001) and from 6.1-12.3 percent as estimated by Patrinos et al. (2006). We assume that:

- a. there is a 6.5 percent rate of return to education (averaging the bottom end of the rates of return reported in the papers above in order to be more conservative in our analysis)
- b. children do not begin to realize the benefits of increased wages until age 18
- c. they do so for 40 years

Under these assumptions, a 0.1-year increase in schooling results in a benefit-cost ratio of 0.65.²¹ Similarly, a 0.7-year increase in schooling results in a benefit-cost ratio of 4.55. Using higher rates of return as assumed in the World Bank Project Appraisal Document (11.2 percent) suggests a correspondingly much higher benefit-cost ratio of 1.12. – 7.84. Thus even the most conservative cost-benefit estimates would suggest that the project did far better than breakeven. This is an underestimate of the benefit given the conservative estimates of returns to education used, the shorter-than-usual time horizon for accrual of benefits as well as the fact that these are only private returns for selected cohorts. Social returns to education have not been factored in, nor have any gains resulting from improved learning.

²¹ The 2012 GDP per capita in PPP terms was US\$4,876. In our calculations of rate of return, we assume that rural wages are a third of this number.

Table 6: Cost-benefit-analysis

	Older cohort		Younger cohort	
	Per beneficiary (\$)	Total (\$)	Per beneficiary (\$)	Total (\$)
Discounted stream of income (B)	96	64,573,510	601	404,598,373
Discounted cost (C)	76	51,469,388	76	51,469,388
B-C	19	13,104,122	525	353,128,985
Return for each USD invested	1.3		7.9	

Assumptions:

Number of beneficiaries = 673,162.

Annual cost per beneficiary = USD 27.

Benefits start at age 18 and continue for 40 years.

Returns to education = 6.5%.

Discount rate = 5%.

Average annual earnings = 33% of 2012 GDP per capita in PPP terms.

Additional years of schooling for younger cohort = 0.7.

Additional years of schooling for older cohort = 0.1.

8. Discussion

This paper showed that expanding access to preschools in rural Indonesia increased enrollment in preschools and improved child development outcomes. Most importantly, some of these positive impacts were sustained into primary school for the younger cohort. The evaluation design of the Indonesia ECED project allowed us to assess treatment effect variation across two cohorts of children who were exposed to key components of the project at different points in the project lifespan. Although the project had positive impacts for both cohorts of children, the fact that these impacts varied across cohorts suggests that the maturity and timing of preschool interventions are important factors to consider in impact evaluations.

There are a number of factors that explain the different outcomes resulting from expanded access to early childhood education services observed across the two cohorts in this evaluation.

The first is that children in the younger cohort were significantly more likely to enroll in project playgroups than the older cohort. The younger cohort was also enrolled in these services for longer than the older cohort. This was likely due both to the older cohort being at the upper end of the appropriate age range for playgroups at the onset of the project, as well as the maturity of the playgroups themselves at the time when the cohorts were able to enroll.

Another factor is how project playgroups evolved between 2009 and 2013 in terms of user fees. Approximately half of the project playgroups were not charging user fees at the beginning of the project, meaning that many of the children in the older cohort accessed the services for free (Brinkman et al. 2015).²² By the end of the project, less than a quarter of the centers were free, with approximately half of all project playgroups charging between 10,000 and 25,000 IDR, which was comparable to the amount charged by non-project playgroups.

Among those children only enrolled in project playgroups, the wealth profile was very different for the two cohorts of children. On average, children in the younger cohort had

²² See Appendix Table 7.

higher wealth z-scores than those in the older cohort.²³ In conjunction with the introduction of fees as the project matured, the change in student composition in our data implies that it was easier for poorer children to enroll in project playgroups early in the project's lifespan than in subsequent years.²⁴

Likewise, the quality of the preschools likely ebbed and flowed during the period under study. Brinkman et al. (2017a) establish the strong link between child development outcomes and the quality of the services being provided, as measured using classroom observation. Emerging evidence on how centers evolved during the life of the project and once project funding ended suggests that quality was not static (Hasan et al., forthcoming). Teacher training was delivered over time. Thus, at the outset of the project, centers began operating without necessarily having a full contingent of trained teachers. This process may have meant that centers had lower quality services during their first year than in later years of the project.

There are some limitations to this study. First, our analysis sample has attrition. Of the 3,089 children who were surveyed at baseline in 2009, 2,894 children (93.69%) were followed up for data collection in 2016. To limit attrition, enumerators were instructed to visit children in their homes if students and their primary caregivers were absent on the day of data collection in schools. As shown in Appendix Table 9, columns 1 and 2, there was no difference in attrition rates between treatment and comparison villages. Moreover, the interaction terms included in column 3 show that the characteristics of those who were not available for follow-up are largely similar across treatment and comparison villages. The table shows that children with less educated mothers and lower baseline gross motor skills were more likely to cease participation in treatment villages compared to similar children in comparison villages. However, overall, we do not find evidence that attrition poses a threat to the validity of our impact estimates.

Another limitation to our study is measurement error of our outcomes. As documented in Appendix Figures 1 and 2, several of our key outcome measures suffered from ceiling effects. This makes it difficult to detect effects that may have existed if an instrument that did not suffer from such effects had been used.

Lastly, we are unable to empirically test *why* the treatment effect varies across the two cohorts. While we posed several plausible mechanisms, we cannot be sure why we found larger effects for the younger cohort than the older cohort at ages 5 and 8.

Despite these limitations, this study has meaningful findings. The results from this study indicate that a low-cost, community-based early childhood program can positively impact child development. For both cohorts, children who resided in treatment villages were more likely to enroll in project playgroups, were enrolled for longer, and had substantially better measures of child development than children in villages where these services were not available. The effects of this exposure persisted into early primary school for the younger cohort, as judged by tests of language, mathematics and reasoning.

There are a number of factors to consider when trying to ensure that the benefits of preschool programs are delivered consistently over time. This is particularly true in low-dose, center-based environments that are expanding in the developing world. As future early childhood education projects are designed and implemented, these myriad considerations will be important to balance against each other if sustained impacts are to be achieved.

²³ See Appendix Table 8.

²⁴ By the time the younger cohort was old enough to enroll in project playgroups, many more playgroups were charging fees. The introduction of fees was a direct response to the project funding coming to an end and the centers needing to devise an alternative sustainability strategy (Hasan et al., forthcoming).

References

- Alatas, Hafid, Brinkman, S., Chang, M.C., Hadiyati, T., Hartono, D., Hasan, A., Hyson, M., Jung, H., Kinnell, A., Pradhan, M.P., and Roesli, R. 2013. Early childhood education and development services in Indonesia. In *Education in Indonesia* (pp. Ch-5). Institute of Southeast Asian Studies.
- Barnett, W.S. 1997. Costs and financing of early child development programs. In *Early child development: Investing in our children's future*, ed. Mary E. Young. Amsterdam: Elsevier.
- Barnett, W.S. 2011. "Effectiveness of Early Educational Intervention." *Science*, 333(6045), 975-78.
- Bhutta, A.T.; M.A. Cleves; P.H. Casey; M.M. Craddock and K.J.S. Anand. 2002. "Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born Preterm: A Meta-Analysis." *Journal of the American Medical Association*, 288(6), 728-37.
- Bouguen, A., Filmer, D., Macours, K and Naudeau, S. 2017. Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment. *The Journal of Human Resources*, February 2017.
- Brennan, L. M.; D. S. Shaw; T. J. Dishion and M. Wilson. 2012. "Longitudinal Predictors of School-Age Academic Achievement: Unique Contributions of Toddler-Age Aggression, Oppositionality, Inattention, and Hyperactivity." *Journal of Abnormal Child Psychology*, 40(8), 1289-300.
- Brinkman, S., A. Hasan, H. Jung, A. Kinnell, M. Pradhan. 2015. The impact of expanding access to early childhood services in rural Indonesia: evidence from two cohorts of children. Policy Research Working Paper, No. WPS 7372, Impact Evaluation series. Washington, D.C. : World Bank Group.
- Brinkman, S., A. Hasan, H. Jung, A. Kinnell, N. Nakajima and M. Pradhan. 2017a. The role of preschool quality in promoting child development: evidence from rural Indonesia, *European Early Childhood Education Research Journal*.
- Brinkman, S., A. Hasan, H. Jung, A. Kinnell, and M. Pradhan. 2017b. The impact of expanding access to early childhood education services in rural Indonesia. *Journal of Labor Economics*. vol. 35, no. S1.
- Brinkman S, Kinnell A, Maika A, Hasan A, Jung H, and Pradhan M. 2016. "Validity and Reliability of the Early Development Instrument in Indonesia." *Child Indicators Research*.
- 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, 140-65.
- Conti, G. and J. Heckman. 2012. "The Economics of Child Well-Being. NBER Working Paper. No. 18466,"
- D'Onise, K.; J.W. Lynch; M. G. Sawyer and R. A. McDermott. 2010. "Can Preschool Improve Child Health Outcomes? A Systematic Review." *Social Science & Medicine*, 70(9), 1423-40.
- Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*. 91(4), September 2001, pp. 795-813.
- Duncan, G.; A. Claessens; A. Huston; L. Pagani; M. Engel; H. Sexton; C. Dowsett; K. Magnusson; P. Klebanov; L. Feinstein, et al. 2007. "School Readiness and Later Achievement." *Developmental Psychology*, 43(6), 1428-46.
- Duncan, G. and K. Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives*, 27(2), 109-32.

- Engle, P.L.; L.C. H. Fernald; H. Alderman; J.R. Behrman; C. O'Gara; A. Yousafzai; M.C. de Mello; M. Hidrobo; N. Ulkuer; I. Ertem, et al. 2011. "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries." *The Lancet*, 378(9799), 1339-53.
- Evans, Judith L., Robert G. Meyers, and Ellen Ilfeld. 2000. *Early childhood counts: A programming guide on early childhood care for development*. Washington, DC: World Bank Publications.
- Feinstein, L. and K. Duckworth. 2006. "Development in the Early Years: Its Importance for School Performance and Adult Outcomes.," *The Wider benefits of Learning Research Report Series No. 20*. London: Institute of Education,
- Hasan, A.; M. Hyson and M. Chu Chang (eds.). 2013. *Early Childhood Education and Development in Poor Villages of Indonesia: Strong Foundations, Later Success* The World Bank.
- Hasan, Amer, Haeil Jung, Angela Kinnell, Amelia Maika, Nozomi Nakajima, Menno Pradhan. Forthcoming. *Built to Last: Sustainability of Early Childhood Education Services in Rural Indonesia*.
- Heckman, J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science*, 312(5782), 1900-02.
- Hertzman, C. 2013. "Commentary on the Symposium: Biological Embedding, Life Course Development, and the Emergence of a New Science1." *Annual Review of Public Health*, 34(1), 1-5.
- Janus, M. and Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): a measure of children's school readiness. *Canadian Journal of Behavioural Science* 39 (1): 1–22.
- Jung, Haeil and Amer Hasan. 2016. The impact of early childhood education on early achievement gaps in Indonesia, *Journal of Development Effectiveness*, 8:2, 216-233,
- Lynch, J.W. and G. Davey-Smith. 2005. "A Life Course Approach to Chronic Disease Epidemiology." *Annual Review of Public Health*, 26(1), 1-35.
- Lynch, J.W.; C. Law; S. Brinkman; C. Chittleborough and M. Sawyer. 2010. "Inequalities in Child Healthy Development: Some Challenges for Effective Implementation." *Social Science and Medicine*, 71(7), 1219-374.
- Martinez, Sebastian, Sophie Naudeau, Vitor Azevedo Pereira. 2017. *Preschool and child development under extreme poverty: evidence from a randomized experiment in rural Mozambique* (English). Policy Research working paper; no. WPS 8290; Impact Evaluation series. Washington, D.C. : World Bank Group.
- Melhuish, E.C.; K. Sylva; P. Sammons; I. Siraj-Blatchford; B. Taggart; M.B. Phan and A. Malin. 2008. "Preschool Influences on Mathematics Achievement." *Science*, 321, 1161-62.
- Moffitt, T.E.; L. Arseneault; D. Belsky; N. Dickson; R.J. Hancox; H. Harrington; R. Houts; R. Poulton; B.W. Roberts; S. Ross, et al. 2011. "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety." *Proceedings of the National Academy of Sciences*, 108(7), 2693-98.
- Moser, S. E.; S. G. West and J. N. Hughes. 2012. "Trajectories of Math and Reading Achievement in Low-Achieving Children in Elementary School: Effects of Early and Later Retention in Grade." *Journal of Educational Psychology*, 104(3), 603-21.
- Nakajima, N., A. Hasan, H. Jung, S. Brinkman, M. Pradhan, A. Kinnell. 2016. *Investing in school readiness: an analysis of the cost-effectiveness of early childhood education pathways in rural Indonesia* . Policy Research Working Paper, No. WPS 7832, WDR 2018 background paper. Washington, D.C.: World Bank Group.

- Nores, M. and S. Barnett. 2010. "Benefits of Early Childhood Interventions across the World: (under) Investing in the Very Young." *Economics of Education Review*, 29(2), 271-82.
- Office of Population Studies. 2005. A Study of the Effects of Early Childhood Interventions on Children's Physiological, Cognitive and Social Development. Cebu City, Philippines: Office of Population Studies, University of San Carlos.
- Patrinos, Harry Anthony, Cris Ridao-Cano and Chris Sakellariou. 2006. Estimating the returns to education: accounting for heterogeneity in ability. Policy Research Working Paper Series No. 4040. Washington, DC: World Bank.
- Pradhan, M.; S. A. Brinkman; A. Beatty; A. Maika; E. Satriawan; J. de Ree and A. Hasan. 2013. "Evaluating a Community-Based Early Childhood Education and Development Program in Indonesia: Study Protocol for a Pragmatic Cluster Randomized Controlled Trial with Supplementary Matched Control Group." *Trials*, 14, 16.
- Sen, Amartya. 1999. Development as Freedom. Oxford University Press.
- Shonkoff, J.P. 2014. "Changing the Narrative for Early Childhood Investment." *Journal of the American Medical Association Pediatrics*, 168(2), 105-06.
- Shonkoff, J.P. and D. Phillips. 2000. From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington, D.C.: National Academy Press.
- Stevens, Katharine B., and Elizabeth English. Does Pre-K Work? The Research on Ten Early Childhood Programs - and What it tells us? American Enterprise Institute.
- Tanner, Jeffery C., Tara Candland, and Whitney S. Odden, 2015. Later Impacts of Early Childhood Interventions: A Systematic Review. IEG Working Paper 2015/3.
- Todd, P. E. and K. I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal*, 113(485), F3-F33.
- Victora, C.G.; L. Adair; C. Fall; P.C. Hallal; R. Martorell; L. Richter and H.S. Sachdev. 2008. "Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital." *The Lancet*, 371(9609), 340-57.
- Zubrick, S., G. J. Smith, J. M. Nicholson, A. V. Sanson, and T. A. Jackiewicz. 2008. Parenting and Families in Australia. Canberra: FaHCSIA (Department of Families, Housing, Community Services and Indigenous Affairs)

Appendix Table 1. Impact on enrollment outcomes for the younger cohort by batch

Outcome		All	Poor	Non-poor	Low Parental Score	High Parental Score		
Ever enrolled in project playgroup until survey year	2013	Batch 1	0.510*** (0.030)	0.517*** (0.041)	0.506*** (0.034)	0.513*** (0.035)	0.510*** (0.036)	
		Batch 2	0.488*** (0.031)	0.472*** (0.040)	0.502*** (0.033)	0.489*** (0.034)	0.488*** (0.037)	
	2016	Batch 1	0.508*** (0.030)	0.562*** (0.040)	0.470*** (0.035)	0.477*** (0.036)	0.544*** (0.036)	
		Batch 2	0.491*** (0.031)	0.520*** (0.039)	0.474*** (0.037)	0.458*** (0.036)	0.524*** (0.034)	
	Ever enrolled in non-project playgroup until survey year	2013	Batch 1	-0.226*** (0.039)	-0.209*** (0.045)	-0.241*** (0.046)	-0.223*** (0.042)	-0.228*** (0.046)
			Batch 2	-0.214*** (0.039)	-0.178*** (0.046)	-0.241*** (0.044)	-0.207*** (0.043)	-0.219*** (0.047)
2016		Batch 1	-0.238*** (0.039)	-0.208*** (0.047)	-0.255*** (0.045)	-0.238*** (0.043)	-0.238*** (0.049)	
		Batch 2	-0.214*** (0.039)	-0.177*** (0.048)	-0.237*** (0.045)	-0.191*** (0.044)	-0.236*** (0.047)	
Ever enrolled in kindergarten until survey year		2013	Batch 1	-0.070 (0.045)	-0.063 (0.058)	-0.090 (0.046)	-0.076 (0.053)	-0.063 (0.049)
			Batch 2	-0.079 (0.046)	-0.084 (0.059)	-0.081 (0.048)	-0.102 (0.054)	-0.056 (0.049)
	2016	Batch 1	-0.083 (0.046)	-0.067 (0.063)	-0.096* (0.043)	-0.116* (0.052)	-0.045 (0.050)	
		Batch 2	-0.092* (0.046)	-0.069 (0.060)	-0.107* (0.046)	-0.087 (0.049)	-0.103 (0.054)	
Months enrolled in project playgroup until survey year	2013	Batch 1	8.336*** (0.595)	8.172*** (0.768)	8.487*** (0.679)	8.184*** (0.666)	8.533*** (0.720)	
		Batch 2	7.267*** (0.515)	6.513*** (0.610)	7.805*** (0.589)	7.353*** (0.583)	7.207*** (0.628)	
	2016	Batch 1	9.338*** (0.614)	9.668*** (0.774)	9.067*** (0.686)	9.380*** (0.713)	9.378*** (0.813)	
		Batch 2	8.130*** (0.543)	8.232*** (0.698)	8.083*** (0.634)	7.963*** (0.607)	8.424*** (0.700)	

Appendix Table 1. Impact on enrollment outcomes for the younger cohort by batch

Outcome		All	Poor	Non-poor	Low Parental Score	High Parental Score		
Months enrolled in non-project playgroup until survey year	2013	Batch 1	-2.990*** (0.677)	-2.151*** (0.616)	-3.686*** (0.903)	-2.631*** (0.627)	-3.466*** (0.897)	
		Batch 2	-3.012*** (0.657)	-1.996** (0.603)	-3.775*** (0.854)	-2.465*** (0.622)	-3.748*** (0.860)	
	2016	Batch 1	-3.378*** (0.676)	-2.700*** (0.714)	-3.856*** (0.863)	-3.211*** (0.696)	-3.604*** (0.918)	
		Batch 2	-3.179*** (0.684)	-2.390*** (0.702)	-3.714*** (0.862)	-2.740*** (0.725)	-3.851*** (0.870)	
	Months enrolled in kindergarten until survey year	2013	Batch 1	-0.118 (0.478)	0.213 (0.554)	-0.534 (0.549)	-0.035 (0.538)	-0.234 (0.602)
			Batch 2	-0.274 (0.516)	-0.021 (0.596)	-0.469 (0.587)	-0.188 (0.545)	-0.423 (0.616)
2016		Batch 1	-1.030 (0.739)	-0.548 (0.959)	-1.396 (0.756)	-1.648* (0.807)	-0.131 (0.884)	
		Batch 2	-1.044 (0.781)	-0.670 (0.997)	-1.154 (0.828)	-1.003 (0.806)	-1.253 (0.979)	

p<0.001***; p<0.01**; p<0.05*

Note: Each cell is the result of a separate regression. Standard errors clustered at the village level are in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 2. Impact on EDI domains for the younger cohort by batch

Outcome	Year	Batch	All	Poor	Non-poor	Low Parenting Score	High Parenting Score	
Physical health and well-being (SD)	2013	Batch 1	0.147* (0.059)	0.063 (0.092)	0.197** (0.070)	0.070 (0.071)	0.242** (0.086)	
		Batch 2	0.265*** (0.057)	0.168 (0.091)	0.331*** (0.065)	0.241*** (0.067)	0.302*** (0.084)	
	2016	Batch 1	-0.080 (0.075)	0.000 (0.117)	-0.137* (0.066)	-0.041 (0.091)	-0.139 (0.079)	
		Batch 2	0.076 (0.067)	0.209* (0.105)	-0.014 (0.063)	0.136 (0.082)	-0.015 (0.070)	
	Social competence (SD)	2013	Batch 1	0.061 (0.059)	0.138 (0.088)	0.001 (0.064)	0.030 (0.063)	0.093 (0.091)
			Batch 2	-0.023 (0.063)	0.030 (0.092)	-0.055 (0.068)	-0.038 (0.065)	-0.011 (0.091)
2016		Batch 1	-0.037 (0.057)	-0.004 (0.078)	-0.056 (0.064)	-0.066 (0.071)	0.005 (0.065)	
		Batch 2	0.011 (0.058)	0.094 (0.077)	-0.040 (0.066)	-0.008 (0.075)	0.040 (0.063)	
Emotional maturity (SD)	2013	Batch 1	0.091 (0.068)	0.206* (0.097)	0.006 (0.069)	0.086 (0.087)	0.096 (0.073)	
		Batch 2	0.138* (0.063)	0.275** (0.093)	0.043 (0.064)	0.139 (0.079)	0.137* (0.067)	
	2016	Batch 1	0.016 (0.050)	0.054 (0.067)	-0.003 (0.064)	0.039 (0.068)	-0.025 (0.065)	
		Batch 2	-0.009 (0.052)	0.134* (0.068)	-0.119 (0.065)	0.003 (0.068)	-0.026 (0.062)	
Language and cognitive development (SD)	2013	Batch 1	0.135* (0.061)	0.157 (0.087)	0.109 (0.066)	0.149* (0.067)	0.120 (0.086)	
		Batch 2	0.015 (0.068)	0.051 (0.100)	-0.010 (0.072)	0.059 (0.077)	-0.038 (0.087)	
	2016	Batch 1	0.060 (0.045)	0.060 (0.076)	0.073 (0.043)	0.090 (0.059)	0.021 (0.050)	
		Batch 2	0.014 (0.044)	0.029 (0.071)	-0.002 (0.051)	0.046 (0.058)	-0.036 (0.052)	
Communication and general knowledge (SD)	2013	Batch 1	-0.012 (0.081)	-0.149 (0.107)	0.093 (0.086)	-0.044 (0.089)	0.038 (0.104)	
		Batch 2	0.001 (0.085)	-0.034 (0.117)	0.023 (0.085)	-0.047 (0.086)	0.068 (0.109)	
	2016	Batch 1	-0.200** (0.072)	-0.244** (0.082)	-0.152 (0.088)	-0.173* (0.086)	-0.249** (0.084)	
		Batch 2	-0.124 (0.070)	-0.080 (0.077)	-0.143 (0.085)	-0.125 (0.083)	-0.126 (0.076)	

p<0.001***; p<0.01**; p<0.05*

Note: Each cell is the result of a separate regression. Standard errors clustered at the village level are in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 3. Impact on primary school test scores for the younger cohort by batch

Outcome	Batch	All	Poor	Non-poor	Low Parenting Score	High Parenting Score
Language - match picture (SD)	Batch 1	0.155* (0.067)	0.058 (0.092)	0.238** (0.074)	0.190** (0.071)	0.093 (0.095)
	Batch 2	0.113 (0.065)	0.005 (0.092)	0.189* (0.075)	0.152* (0.070)	0.047 (0.093)
Language - mention objects (SD)	Batch 1	0.003 (0.063)	-0.078 (0.087)	0.080 (0.072)	-0.015 (0.070)	0.039 (0.079)
	Batch 2	-0.054 (0.058)	-0.057 (0.072)	-0.047 (0.071)	-0.048 (0.065)	-0.070 (0.076)
Math - summation (SD)	Batch 1	0.125* (0.060)	0.093 (0.088)	0.157* (0.064)	0.130 (0.068)	0.125 (0.087)
	Batch 2	0.016 (0.061)	0.024 (0.086)	-0.000 (0.071)	0.022 (0.070)	0.015 (0.080)
Math - order numbers (SD)	Batch 1	0.173** (0.065)	0.167 (0.092)	0.189** (0.071)	0.110 (0.073)	0.267** (0.087)
	Batch 2	0.081 (0.065)	0.048 (0.093)	0.105 (0.076)	0.042 (0.072)	0.141 (0.085)
Raven's matrices (SD)	Batch 1	0.007 (0.052)	-0.000 (0.075)	0.017 (0.064)	0.021 (0.062)	-0.014 (0.075)
	Batch 2	-0.046 (0.051)	-0.038 (0.075)	-0.058 (0.064)	-0.042 (0.061)	-0.053 (0.078)

p<0.001***; p<0.01**; p<0.05*

Note: Each cell is the result of a separate regression. Standard errors clustered at the village level are in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 4. Impact heterogeneity by wealth & parenting scores on enrollment outcomes for the younger cohort

Outcome	Year		Household wealth	Parental score
			Non-poor (0) vs. Poor (1)	Low (0) vs. High (1)
Ever enrolled in project playgroup until survey year	2013	Treatment Interaction	-0.018	0.007
		(S.E.)	(0.029)	(0.025)
		Obs.	2,778	2,778
	2016	Treatment Interaction	0.065*	-0.073*
		(S.E.)	(0.032)	(0.029)
		Obs.	2,894	2,894
Ever enrolled in non-project playgroup until survey year	2013	Treatment Interaction	0.053	0.007
		(S.E.)	(0.044)	(0.038)
		Obs.	2,778	2,778
	2016	Treatment Interaction	0.050	0.024
		(S.E.)	(0.046)	(0.043)
		Obs.	2,894	2,894
Ever enrolled in kindergarten until survey year	2013	Treatment Interaction	0.015	-0.035
		(S.E.)	(0.048)	(0.040)
		Obs.	2,778	2,778
	2016	Treatment Interaction	0.035	-0.021
		(S.E.)	(0.045)	(0.043)
		Obs.	2,894	2,894
Months enrolled in project playgroup until survey year	2013	Treatment Interaction	-0.945	0.002
		(S.E.)	(0.511)	(0.463)
		Obs.	2,778	2,778
	2016	Treatment Interaction	0.226	-0.763
		(S.E.)	(0.553)	(0.525)
		Obs.	2,894	2,894
Months enrolled in non-project playgroup until survey year	2013	Treatment Interaction	1.686*	0.480
		(S.E.)	(0.768)	(0.644)
		Obs.	2,778	2,778
	2016	Treatment Interaction	1.166	0.556
		(S.E.)	(0.791)	(0.713)
		Obs.	2,894	2,894
Months enrolled in kindergarten until survey year	2013	Treatment Interaction	0.823	-0.598
		(S.E.)	(0.721)	(0.699)
		Obs.	2,778	2,778
	2016	Treatment Interaction	0.659	0.107
		(S.E.)	(0.510)	(0.466)
		Obs.	2,894	2,894

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at village level in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 5. Impact heterogeneity on EDI domains for the younger cohort by wealth & parenting scores

Outcome	Year		Household wealth	Parenting score
			Non-poor (0) vs. Poor (1)	Low (0) vs. High (1)
Physical health and well-being (SD)	2013	Treatment interaction	-0.141	-0.108
		(S.E.)	(0.096)	(0.090)
		Obs.	2,770	2,770
	2016	Treatment interaction	0.159	-0.018
		(S.E.)	(0.100)	(0.079)
		Obs.	2,877	2,877
Social competence (SD)	2013	Treatment interaction	0.106	0.001
		(S.E.)	(0.089)	(0.084)
		Obs.	2,769	2,769
	2016	Treatment interaction	0.079	-0.080
		(S.E.)	(0.075)	(0.078)
		Obs.	2,877	2,877
Emotional maturity (SD)	2013	Treatment interaction	0.224**	-0.005
		(S.E.)	(0.083)	(0.076)
		Obs.	2,770	2,770
	2016	Treatment interaction	0.153*	-0.070
		(S.E.)	(0.076)	(0.081)
		Obs.	2,877	2,877
Language and cognitive development (SD)	2013	Treatment interaction	0.060	0.047
		(S.E.)	(0.088)	(0.079)
		Obs.	2,770	2,770
	2016	Treatment interaction	0.005	0.007
		(S.E.)	(0.087)	(0.089)
		Obs.	2,877	2,877
Communication and general knowledge (SD)	2013	Treatment interaction	-0.148	-0.127
		(S.E.)	(0.091)	(0.085)
		Obs.	2,771	2,771
	2016	Treatment interaction	-0.019	-0.036
		(S.E.)	(0.080)	(0.081)
		Obs.	2,877	2,877

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at village level in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 6 Impact heterogeneity by wealth & parental scores on test scores for the younger cohort

Outcome		Household wealth	Parenting score
		Non-poor (0) vs. Poor (1)	Low (0) vs. High (1)
Language - match picture (SD)	Treatment interaction	-0.171	0.105
	(S.E.)	(0.093)	(0.082)
	Obs.	2,862	2,862
Language - mention objects (SD)	Treatment interaction	-0.059	-0.003
	(S.E.)	(0.080)	(0.065)
	Obs.	2,862	2,862
Math - summation (SD)	Treatment interaction	0.005	-0.036
	(S.E.)	(0.083)	(0.078)
	Obs.	2,862	2,862
Math - order numbers (SD)	Treatment interaction	-0.028	-0.108
	(S.E.)	(0.092)	(0.080)
	Obs.	2,862	2,862
Raven's matrices (SD)	Treatment interaction	0.034	0.011
	(S.E.)	(0.079)	(0.076)
	Obs.	2,862	2,862

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at the village level are in parentheses. All regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language).

Appendix Table 7. Percentage of centers charging monthly fees, by year and type of center

	Project playgroups			Non-project playgroups	
	2010	2013	2016	2013	2016
Percent of centers charging					
No fees	52	30	22	19	17
Less than 10,000 IDR	38	33	20	19	11
10,000 - 25,000 IDR	9	31	48	52	45
25,000 - 50,000 IDR	1	4	8	7	22
More than 50,000 IDR	0	2	2	2	5

Note: All numbers are percent of centers surveyed. Not necessarily the same centers in each survey round but information is from the same villages over time. Non-project centers were not surveyed in 2010. 1 USD = approximately 10,000 IDR in 2010.

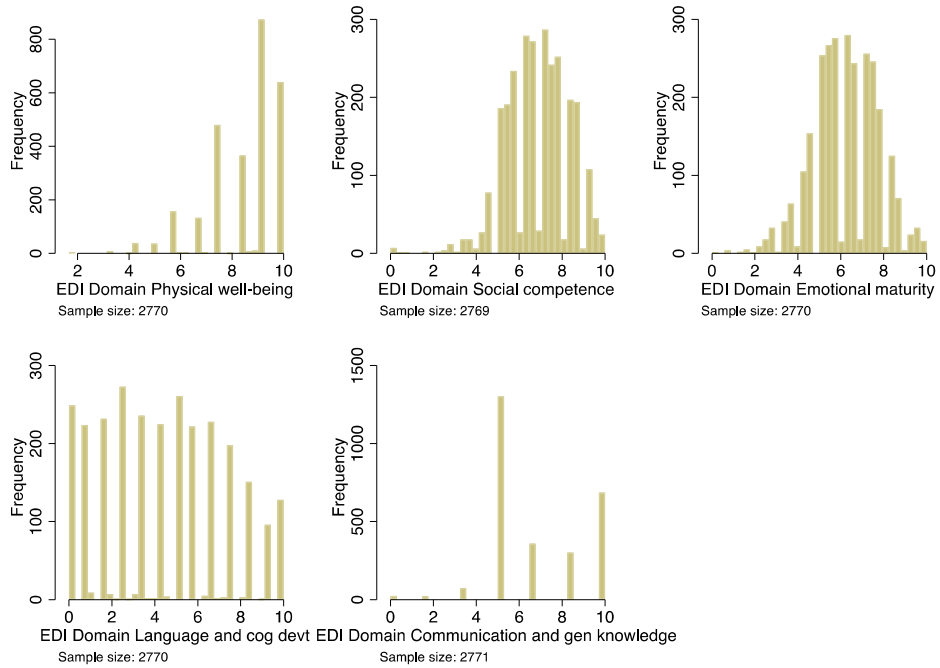
Appendix Table 8. Wealth profile and mother's education for the two cohorts at age 8

	Mother's education primary school or less (1=Yes)		Wealth z-score	
	Younger cohort	Older cohort	Younger cohort	Older cohort
No ECED	0.767	0.761	-0.410	-0.538
	<i>0.423</i>	<i>0.427</i>	<i>1.100</i>	1.085
	322	452	322	452
Project playgroup only	0.575	0.650	-0.029	-0.175
	<i>0.496</i>	<i>0.478</i>	<i>0.784</i>	0.919
	179	223	179	223
Non-project playgroup only	0.552	0.669	-0.089	-0.175
	0.498	0.472	1.04	0.969
	420	130	420	130
Kindergarten only	0.483	0.467	0.161	0.224
	<i>0.499</i>	<i>0.499</i>	<i>0.919</i>	<i>0.896</i>
	878	841	878	841
Project playgroup & other services	0.483	0.458	0.092	0.377
	<i>0.500</i>	<i>0.500</i>	<i>0.969</i>	0.691
	775	120	775	120
Non-project playgroup & other services	0.426	0.439	0.177	0.411
	<i>0.495</i>	<i>0.498</i>	<i>0.941</i>	<i>0.807</i>
	540	187	540	187

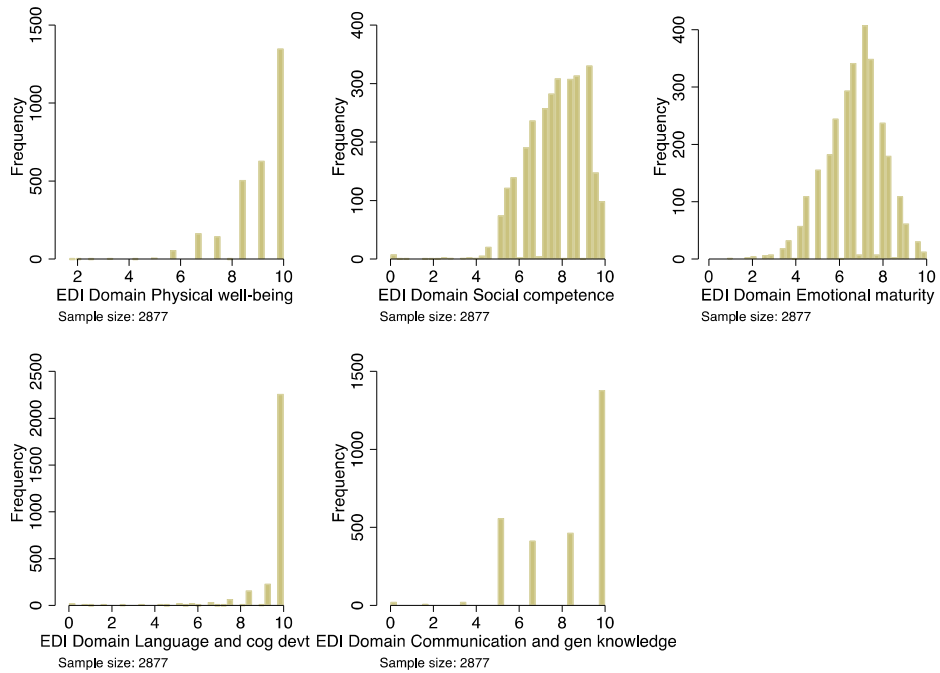
Notes: Standard deviations are in italics. Sample size is in bold; consistent with our empirical strategies, 1 year-old cohort includes batches 1, 3 and 5 whereas 4 year-old cohort includes only batch 3 and 5. Sample means are reported for children in each cohort who report at least some months of attendance of various types of services and those who report zero months of attendance in ECED (for no ECED category).

Appendix Figure 1. Distribution of raw EDI domain scores for the younger cohort in 2013 and 2016

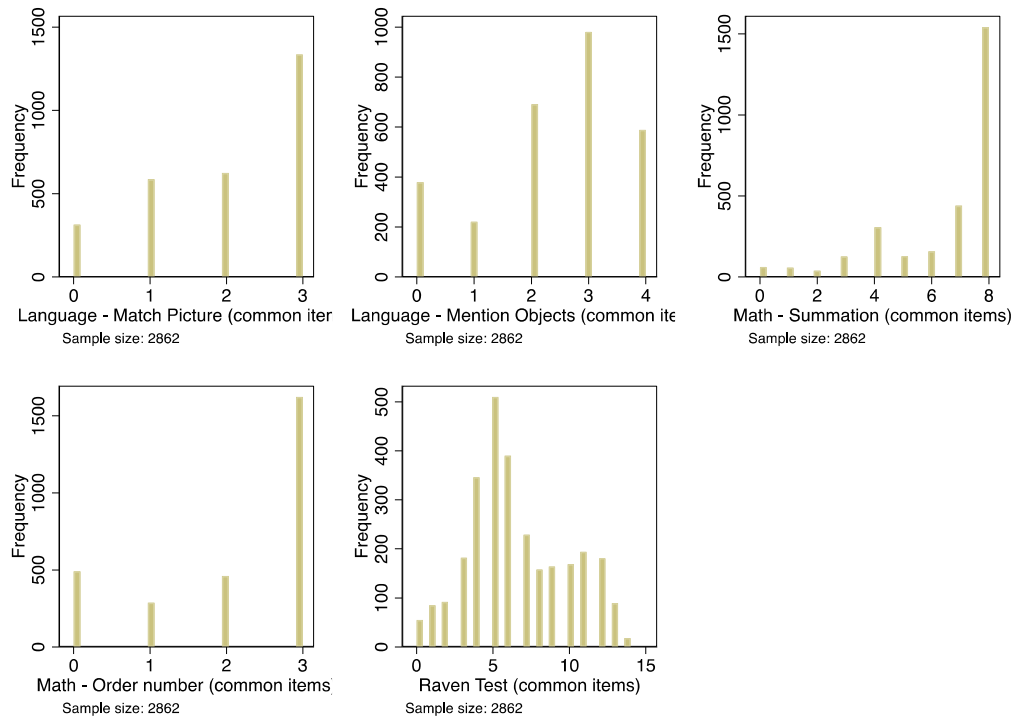
Panel A: 2013 (age 5)



Panel B: 2016 (age 8)



Appendix Figure 2. Distribution of raw test scores for the younger cohort in 2016



Appendix Table 9. Attrition rates for the younger cohort by village

	Attrition (1 = Yes) (1)	Attrition (1 = Yes) (2)	Attrition (1 = Yes) (3)
Batch 1	-0.007 (0.012)		
Batch 2	-0.004 (0.012)		
Treatment		-0.005 (0.010)	0.057 (0.094)
Baseline BMI	0.004 (0.002)	0.004 (0.002)	0.006 (0.004)
Baseline cognitive skills	0.001 (0.004)	0.001 (0.004)	-0.002 (0.006)
Baseline fine motor skills	-0.002 (0.002)	-0.002 (0.002)	0.000 (0.004)
Baseline gross motor skills	-0.000 (0.001)	-0.000 (0.001)	0.004 (0.002)
Baseline language skills	0.004 (0.004)	0.004 (0.004)	-0.003 (0.008)
Child age	-0.018 (0.019)	-0.019 (0.019)	-0.050 (0.034)
Household size	0.008* (0.003)	0.008* (0.003)	0.009 (0.007)
Wealth z-score	-0.001 (0.005)	-0.001 (0.005)	-0.007 (0.009)
Parenting z-score	0.008 (0.005)	0.008 (0.005)	-0.001 (0.009)
Mother education	-0.020 (0.010)	-0.020 (0.010)	0.015 (0.020)
Male (1=Yes)	0.012 (0.009)	0.012 (0.009)	-0.005 (0.014)
Baseline BMI x Treatment			-0.004 (0.005)
Baseline cognitive skills x Treatment			0.004 (0.007)
Baseline fine motor skills x Treatment			-0.003 (0.005)
Baseline gross motor skills x Treatment			-0.006* (0.003)
Baseline language skills x Treatment			0.010 (0.009)
Child age x Treatment			0.046 (0.041)
Household size x Treatment			-0.002 (0.008)
Wealth z-score x Treatment			0.010 (0.011)
Parenting z-score x Treatment			0.014 (0.011)
Mother education x Treatment			-0.050* (0.023)
Male (1=Yes) x Treatment			0.025 (0.018)
Constant	0.009 (0.045)	0.010 (0.046)	-0.033 (0.075)
Observations	3,089	3,089	3,089

*** p<0.001, ** p<0.01, * p<0.05

Note: Each column is the result of a separate regression. Standard errors clustered at the village level are in parentheses.

Appendix B – Identification strategy for 4-year-old cohort

For the 4 year-old the point estimates reported in Table 5 are obtained using a difference-in-difference approach. The difference-in-difference model is estimated as follows:

$$Y_{ijt} = \varphi_i + \sum_{t=1}^2 D_t \tau_t + Batch\ 2_j \alpha + \sum_{t=1}^2 Batch\ 2_j D_t \delta_t + X_{ijt} \beta + u_{ijt} \quad (2)$$

where $Batch\ 2_j$ is a dummy variable which takes on a value of 1 for villages treated in 2010, and a value of 0 for comparison villages. Thus, α indicates the baseline ($t = 0$) difference between the two groups. φ_i is a child fixed effect. D_t is the time dummy for $t = 1$ (2010 or midline) or 2 (2013 or endline), which controls for age and time effects in the model. The coefficient δ_t at $t = 1$ or 2 is the difference-in-difference estimator at midline and endline, respectively.

Appendix C - Supplementary Table 1. Impact on EDI of the 1 year-old cohort (Long caregiver EDI)

Outcome	Survey Year		All	Poor	Non-poor	Low Parental Score	High Parental Score
Physical health and well-being (SD)	2013	Coeff.	0.210***	0.133	0.261***	0.160*	0.268***
		(S.E.)	(0.051)	(0.088)	(0.058)	(0.066)	(0.070)
		Comp. mean	-0.151	-0.159	-0.145	-0.198	-0.0983
	Obs.	2,770	1,194	1,576	1,533	1,237	
	2016	Coeff.	-0.043	0.004	-0.078	-0.053	-0.041
		(S.E.)	(0.064)	(0.097)	(0.064)	(0.083)	(0.070)
Comp. mean		0.0554	-0.0755	0.152	-0.146	0.268	
Obs.	2,877	1,279	1,598	1,823	1,054		
Social competence (SD)	2013	Coeff.	0.030	0.084	-0.014	0.038	0.013
		(S.E.)	(0.058)	(0.087)	(0.060)	(0.061)	(0.085)
		Comp. mean	-0.0128	-0.186	0.111	-0.191	0.185
	Obs.	2,769	1,192	1,577	1,534	1,235	
	2016	Coeff.	0.009	0.044	-0.015	-0.029	0.033
		(S.E.)	(0.048)	(0.066)	(0.057)	(0.068)	(0.061)
Comp. mean		0.0251	-0.0294	0.0655	-0.293	0.361	
Obs.	2,877	1,279	1,598	1,823	1,054		
Emotional maturity (SD)	2013	Coeff.	0.094	0.181*	0.034	0.073	0.111
		(S.E.)	(0.055)	(0.081)	(0.058)	(0.069)	(0.063)
		Comp. mean	-0.0934	-0.207	-0.0127	-0.370	0.212
	Obs.	2,770	1,193	1,577	1,534	1,236	
	2016	Coeff.	0.028	0.102	-0.028	-0.006	0.048
		(S.E.)	(0.041)	(0.056)	(0.054)	(0.059)	(0.059)
Comp. mean		-0.00461	-0.139	0.0946	-0.350	0.359	
Obs.	2,877	1,279	1,598	1,823	1,054		
Language and cognitive development (SD)	2013	Coeff.	0.065	0.079	0.043	0.091	0.044
		(S.E.)	(0.064)	(0.092)	(0.067)	(0.069)	(0.080)
		Comp. mean	-0.0427	-0.279	0.127	-0.145	0.0705
	Obs.	2,770	1,193	1,577	1,534	1,236	
	2016	Coeff.	0.071	0.079	0.073	0.059	0.078
		(S.E.)	(0.048)	(0.075)	(0.055)	(0.066)	(0.063)
Comp. mean		-0.0287	-0.119	0.0381	-0.0901	0.0361	
Obs.	2,877	1,279	1,598	1,823	1,054		
Communication skills and general knowledge (SD)	2013	Coeff.	0.008	-0.046	0.042	-0.050	0.077
		(S.E.)	(0.077)	(0.107)	(0.078)	(0.081)	(0.098)
		Comp. mean	0.00769	-0.0962	0.0821	-0.0130	0.0307
	Obs.	2,771	1,194	1,577	1,534	1,237	
	2016	Coeff.	-0.113	-0.119	-0.100	-0.139	-0.096
		(S.E.)	(0.063)	(0.071)	(0.080)	(0.082)	(0.067)
Comp. mean		0.115	0.105	0.123	-0.0694	0.310	
Obs.	2,877	1,279	1,598	1,823	1,054		

p<0.001***; p<0.01**; p<0.05*

Note: Each cell-block is the result of a separate regression. Standard errors clustered at village level in parentheses. “Comp. mean” refers to the comparison group mean for the outcome variable. Column (1) regressions control for child characteristics (age, household size, household wealth asset index, parenting practices, mother's education and gender) and baseline child development measures (BMI, cognitive, fine motor, gross motor, and language). Columns (2) and (3) regressions use the same controls as column (1) except they exclude household wealth. Columns (4) and (5) regressions use the same controls as column (1) except they exclude parenting practices.