Contrasting Experiences

Hasan, Amer; Jung, Haeil; Kinnell, Angela; Maika, Amelia; Nakajima, Nozomi; Pradhan, Menno

published in
Journal of Research on Educational Effectiveness
2021

DOI (link to publisher)
10.1080/19345747.2020.1839989

document version
Publisher's PDF, also known as Version of record

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

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

Amer Hasan, Haeil Jung, Angela Kinnell, Amelia Maika, Nozomi Nakajima & Menno Pradhan

To cite this article: Amer Hasan, Haeil Jung, Angela Kinnell, Amelia Maika, Nozomi Nakajima & Menno Pradhan (2021) Contrasting Experiences: Understanding the Longer-Term Impact of Improving Access to Pre-Primary Education in Rural Indonesia, Journal of Research on Educational Effectiveness, 14:1, 28-56, DOI: 10.1080/19345747.2020.1839989

To link to this article: https://doi.org/10.1080/19345747.2020.1839989

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2021, VOL. 14, NO. 1, 28–56
https://doi.org/10.1080/19345747.2020.1839989

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

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

Amer Hasan[a], Haeil Jung[b], Angela Kinnell[c], Amelia Maika[d], Nozomi Nakajima[e], and Menno Pradhan[f]

[a]Education Global Practice, World Bank, Washington, DC, USA; [b]Department of Public Administration, Korea University, Seoul, Republic of Korea; [c]School of Psychology, University of Adelaide, Adelaide, Australia; [d]Department of Sociology, Gadjah Mada University, Yogyakarta, Indonesia; [e]Harvard Graduate School of Education, Harvard University, Cambridge, Massachusetts, USA; [f]University of Amsterdam, Vrije Universiteit Amsterdam and Tinbergen Institute, Amsterdam, The Netherlands

ABSTRACT

This paper examines the child development outcomes of two cohorts of children who were exposed to the same intervention at different points in time. One cohort was eligible to access play-groups 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 of a five-year project cycle, beginning at age three. The younger cohort was more likely to be exposed to play-groups for longer and at more 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. Moreover, the younger cohort had substantially higher test scores during the early grades of primary school, relative to the older cohort. We document the extent to which program impacts can vary as a result of differences in project implementation.

ARTICLE HISTORY
Received 9 December 2019
Revised 18 July 2020
Accepted 14 September 2020

KEYWORDS
Early childhood education; child development; treatment impacts; differential effects

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 & 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 & Heckman, 2012; Sen, 1999). Many of the problems arising in early childhood have social and financial costs

CONTACT Haeil Jung hijk@korea.ac.kr Department of Public Administration, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea

that cumulatively represent a considerable drain on a country's resources (Feinstein & Duckworth, 2006; Victora et al., 2008).

High-quality pre-primary programs provide an opportunity to mitigate the risk factors that many young children face (Barnett, 2011; Duncan et al., 2007; Duncan & Magnuson, 2013; Heckman, 2006; Shonkoff & 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 pre-primary programs is likely to be large for these children. However, much of the evidence based on the long-term impacts of pre-primary education 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 then that match either the fidelity or the intensity of these interventions.

In developing country settings, rigorous evaluations of pre-primary education programs have emerged in the last decade. Results have ranged from no effect in Cambodia (Bouguen et al., 2018) to positive effects in a variety of settings including Indonesia, (Brinkman et al., 2017; Jung & Hasan, 2016;) Vietnam (as reported in Burger, 2010), Mozambique (Martinez et al., 2017) and Ghana (Wolf et al., 2019) to name but a few. Systematic reviews of pre-primary education interventions in developing countries reinforce the wide range of impacts seen in international settings (Tanner et al., 2015; Nores & Barnett, 2010). The contrasting evidence from different settings has led some to question whether early childhood education can even have lasting impacts (Stevens & English, 2016).

A theme that continues to be much debated is the role of dosage and timing of pre-primary education. Wasik and Snell (2019) provide a synthesis of the evidence on preschool dosage from various settings in the US and note that the preponderance of evidence suggests that more participation in center-based preschool is associated with increased kindergarten readiness—especially for low income children. Nakajima et al. (2019) examine different degrees of exposure to playgroups in rural Indonesia as well as the timing of that exposure and find that longer exposure at the appropriate age leads to better child development outcomes. Tanner et al. (2015) conduct a large-scale meta-analysis and document that estimates from four preschool programs in four different countries—Chile, Colombia, Mozambique, and Uruguay—indicate no clear conclusion as to whether children's subsequent schooling outcomes benefit from larger doses of preschool (Attanasio & Vera-Hernández, 2004; Berlinski et al., 2008; Cortázar Valdés, 2011; Martinez et al., 2012).

In this paper, we contribute to the growing literature on the longer-term impacts of pre-primary education in developing countries as well as the literature on the role of dosage and timing of pre-primary education. We examine the longer-term impacts of early childhood education in rural Indonesia by focusing on children who benefited from increased access to playgroups. The villages studied in this paper received the Indonesia Early Childhood Education and Development (ECED) Project—a project which lasted five years and expanded access to community-based playgroups. The project provided block grants to villages to establish up to two playgroups, providing teacher training, and raising community awareness about the importance of early education.

We analyze data on two cohorts of children: an *older cohort* that was age 4 when the project began and a *younger cohort* that was age 1 when the project began. Both cohorts of children live in the same villages and receive the same project. However, the older and younger cohorts have two important distinctions between them: (1) the phase of project implementation when they were exposed to playgroups and (2) the length of exposure to playgroups at a developmentally appropriate age. In this paper, 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 of the program's impact on these two cohorts.

The rest of the paper proceeds as follows. Section two provides a brief background about the landscape of early childhood education in Indonesia. Next, we introduce this paper's analytical approach and its contribution to the literature. Section four describes the Indonesia Early Childhood Education and Development (ECED) Project. Section five describes the data we use in our analysis. Section six describes the empirical strategy. The seventh section provides the empirical results and the eighth section compares the impacts on the younger cohort and older cohorts. The ninth 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 tenth and final section concludes with a discussion of the findings and their implications for future work.

Early Childhood Education in Indonesia

A variety of different programs for pre-primary education exist in Indonesia and are overseen by different ministries. Two types of pre-primary education programs are dominant: playgroups and kindergartens.

Playgroups

The Ministry of Education and Culture regulates playgroups (*Kelompok Bermain*, KB). These typically are services meant for children ages 3–4 and meet three days per week for two to three hours each day. Playgroups are characterized as play-based learning environments with a combination of both unstructured and structured play activities, typically facilitated by teachers who have nominal formal early childhood education training. Structured play activities generally include songs and dance, and exposure to paints/pencils and paper, and reading sessions where the teacher reads books to the children introducing them to books, letters and numbers. These community playgroups will often have anywhere between 10 and 40 children in some instances.

Kindergartens

In contrast, kindergartens are regulated by either the Ministry of Education and Culture (for *Taman Kanak-kanak*, TK) or the Ministry of Religious Affairs (for *Raudhotul Atfal*, RA). Kindergartens are typically meant for children ages 5–6 and meet five to six days a week for three hours each day. Compared to playgroups, kindergartens emphasize a more academic and structured approach to learning. In addition, the tuition fee for

kindergartens is usually higher than playgroups. While playgroups are typically appropriate for children between the ages of 3 and 4, kindergartens are intended for children between the ages of 5 and 6, and the formal age to start primary school is age 7. However, families often do not adhere to these age limits. Children are often enrolled in playgroups and/or kindergartens at a variety of ages before entering primary school.

Analytical Approach

Given this landscape, we analyze data on two cohorts of children that were selected when the project began—an *older cohort* that was age 4 and a *younger cohort* that was age 1. The older and younger cohorts have two important distinctions between them: (1) the phase of project implementation when they were exposed to playgroups and (2) the length of exposure to playgroups at an *appropriate* age. These distinctions have important implications both for the analysis and the interpretation of our findings. On the first distinction, the older cohort was eligible for playgroups during the project's first year since they were already age 4 when the project started. In contrast, the younger cohort was only eligible for playgroups during the project's third year when they turned 3 years old. As a result, these two cohorts experienced the project at *different* phases of implementation—with important implications for the dose they each received a few years apart. 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 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.[1]

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. (2017), which analyzed the impact of the project on the older cohort at ages 5 and 8. In this paper, we conduct three new analyses. First, we present child development outcomes for the younger cohort also measured at ages 5 and 8. Second, we present new results for the older cohort using primary school test scores collected when children were age 8 that have not been previously published. Third, we contrast the longer-term impacts experienced by the two cohorts.

[1]During this period, the project team also worked with the government to develop a number of policies and guidelines for ECED services at the central and district levels. However, most of these were not formally promulgated until the end of the project in 2013 and do not confound the analyses presented here.

This paper contributes to research on the generalizability of impact evaluations in early childhood education. Pre-primary education 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 pre-primary education 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 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. Specifically, 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 can vary as a result of differences in project implementation. While such variations in impact likely exist in many other settings, we are not aware of other papers that are able to document the implications of these variations.

Indonesia Early Childhood Education and Development (ECED) Project

The Indonesia ECED project was implemented in relatively poor villages in rural Indonesia. Of the 442 districts in the country at the time, 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 using a scoring formula 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]

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.[3]

Second, block grants were provided to each village, in the amount of USD 18,000 per village over three years.[4] Villages could use the grant to establish or support up to two early childhood education centers. No more than 20% of the grant could be spent on new infrastructure. The most common form of services established were playgroups,

[2]The 60 priority villages per district selected to participate in the project were the ones with the highest score using this formula. Proposals were expected for each of the two centers being proposed in a given village. Funding was assured by virtue of inclusion of the village in the project. As per the Project Implementation Completion and Results Report, 99.8% of the expected proposals were received and approved.

[3]The Indonesia Early Childhood Education and Development project was funded through a credit from the International Development Association in the amount of $67.5 million and a grant from the government of the Kingdom of the Netherlands in the amount of $25.3 million. In addition the government of Indonesia provided $34.94 million in funding for the project.

[4]The exchange rate to the dollar was around 9000 IDR to 1 dollar at the time the block grants were given. So USD 18.000 was about IDR 162 million. The minimum wage around that period was about IDR 12 million per year (Online at https://www.bps.go.id/site/resultTab). GDP per capita was about IDR 32 million per year (Source: World Development Indicators). Civil servant teachers earned about IDR 32 million Rupiah per year (Source: Authors calculations based on survey data among 2,700 civil servant teachers collected in Nov 2009).

which are early childhood education services intended to cater to children ages 3 and 4 before they enroll in kindergarten at ages 5 and 6.

Third, the project included a component that provided 200 hours of teacher training for up to two teachers per project playgroup. Teachers were predominantly women from the village, who often had children of their own.[5] Some had prior work experience in health and education. Others had no such prior experience.

Villages were asked to identify community members who would be prepared to lead a playgroup program for children, implement parenting meetings, home visiting, or other informal parent-child programs as needed. Villages used objective criteria to identify potential candidates. Although everyone agreed that greater amounts of formal education were desirable, in response to the realities of typical education levels in project villages, a minimum requirement of secondary school completion (SMA) was set. Additional criteria were interest in young children and commitment to ECED. Because of a lack of local opportunities, few candidates had any prior experience in ECED services, although some were *Posyandu* (local health clinic) volunteers.

The Treatment

The treatment evaluated in this impact evaluation ultimately refers to this package of interventions provided by the ECED project—a community facilitator, block grants to establish playgroups, and teacher training (Hasan et al., 2013).

However, before turning to the evaluation design it is important to describe the playgroups and how they operated. The majority of playgroups included in the evaluation operate 3 or more days per week and meet for at least 2 hours a day. Although playgroups do not use a specific mandated curriculum, most use the government's Generic Menu, which follows the principles of the Beyond Centers and Circle Time (BCCT) curriculum. This play-based methodology, using learning centers to promote holistic development, is introduced to teachers during their training. The essential principles are: (1) children learn through play and social relationships; (2) toys and other learning materials should be concrete (hands-on) and, when possible, locally made; (3) teachers "scaffold" children's learning by being involved in their activities but do not teach in a didactic way; and (4) the day should include a balance of quiet and active and child-initiated and teacher-guided activities that support all areas of child development.

Typically, the day begins with welcoming the children and with songs or movement activities. Next, the teacher may introduce and discuss with the class a topic or theme based on recommendations in the Generic Menu. The topic would be something familiar and interesting to the children and would be reflected in other activities during the day and week. A substantial part of the day is devoted to center time, during which small groups of children choose to play in different areas such as blocks, dramatic (make believe) play, creative arts, puzzles, or other fine motor activities. With the teacher's planning and guidance, during this time children may learn concepts and skills that are important for their holistic development and school readiness. Outdoor play, both teacher-planned games and free play, is also part of a typical day.[6]

[5]Teachers were paid IDR 250,000 per month during the life of the project.
[6]See Hasan et al. (2013) for the playgroup schedule for a typical day.

According to project documents, the cost per child for the project's package of interventions was about US$27 per year.[7] 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.[8]

Evaluation Design

While the project was implemented in 3,000 villages, this paper is based on data for 310 villages. These 310 villages are spread out over nine districts that participated in the ECED project. The districts were selected on the basis of their willingness to cooperate with a randomized rollout of the program and their location, the latter to ensure that the study locations encompassed the regional variety of the project locations.[9] In each of these districts, we sampled three groups of villages: randomly sampled villages which were assigned by lottery to receive the project either in the first round or were assigned to receive the project later. In addition, we sampled a matched comparison group of villages, which were recommended by local administrators as villages that were similar to the randomized villages, but which were not going to receive the project.

This resulted in 218 treatment villages and 92 matched 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.[10] 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.[11] 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. (2017) 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

[7]This is not the fee that households had to pay. The median monthly user fee was 5,000 IDR in 2010 and 10,000 IDR in 2013. Rural households at the time reported a monthly wage of 1.7 million rupiah as per Bureau of Labor Statistics (2016).

[8]See for instance the estimates quoted in Barnett (1997) and Evans et al. (2000).

[9]Districts included in the project were selected on the basis of a formula described in Appendix 3 of Hasan et al. (2013). Districts included in the evaluation were willing to randomize participation. However, within districts villages were selected on the basis of a formula described in Appendix 4. Appendix 7, 8 and 9 test the evaluation design, the internal validity and external validity of the design. These show no differences that would confound the analysis.

[10]Appendix 7–9 in Hasan et al. (2013) document that these villages are well balanced on a range of observable characteristics.

[11]Only 32 children from the two cohorts are siblings to each other.

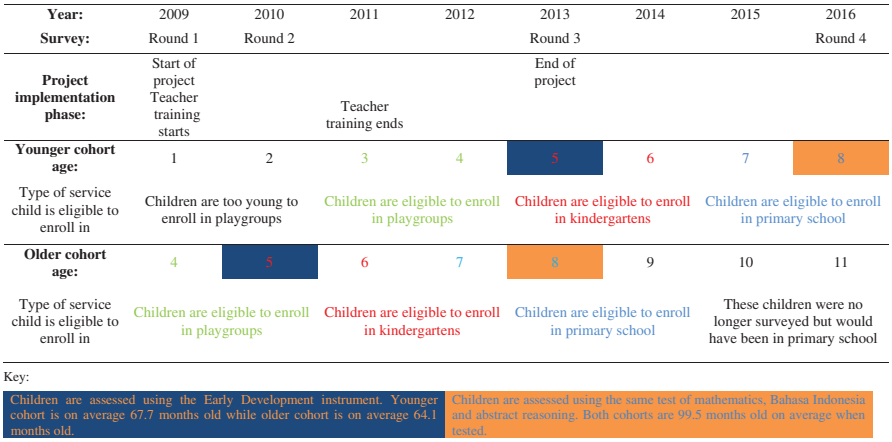

Figure 1. Age of cohort, eligibility for various early childhood education, and project implementation phase for each survey round. Note: Figure depicts ages of the two cohorts studied and what types of services they are eligible for at each age. Younger cohort was surveyed in 2009, 2010, 2013 and 2016. Older cohort was surveyed in 2009, 2010, and 2013.

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 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 first reports 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. We estimate the project impact for the younger cohort at ages 5 and 8 in terms of enrollment in pre-primary education and child development outcomes. It then compares these impacts to those of the older cohort who were eligible to enroll in playgroups early in the project's implementation, in 2009 and 2010. We present existing estimates on a range of outcomes (Brinkman et al., 2017) as well as unpublished estimates on test scores in language (Bahasa Indonesia), mathematics and abstract reasoning at age 8. 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 of the program's impact on these two cohorts.

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.[12]

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 & 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., 2017). In this paper, we use the short-form of the caregiver-rated EDI.[13] Figure 2 Panel A and Panel B show the distributions of EDI measures at age 5 and age 8, respectively. Each domain is scored on a scale of 0 (lowest) to 10 (highest). At age 5, the EDI domains are generally normally distributed, with mean scores ranging between 4.399 and 6.852. The exception is the physical health and well-being domain, which is left skewed with a mean of 8.48. At age 8, the EDI domains are all left skewed, with the exception of the emotional maturity domain. These descriptive figures suggest ceiling effects are present with the EDI domain at age 8, which may contribute to measurement error. For our analysis, 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

[12]Data were collected in 2013 and again in 2016.

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

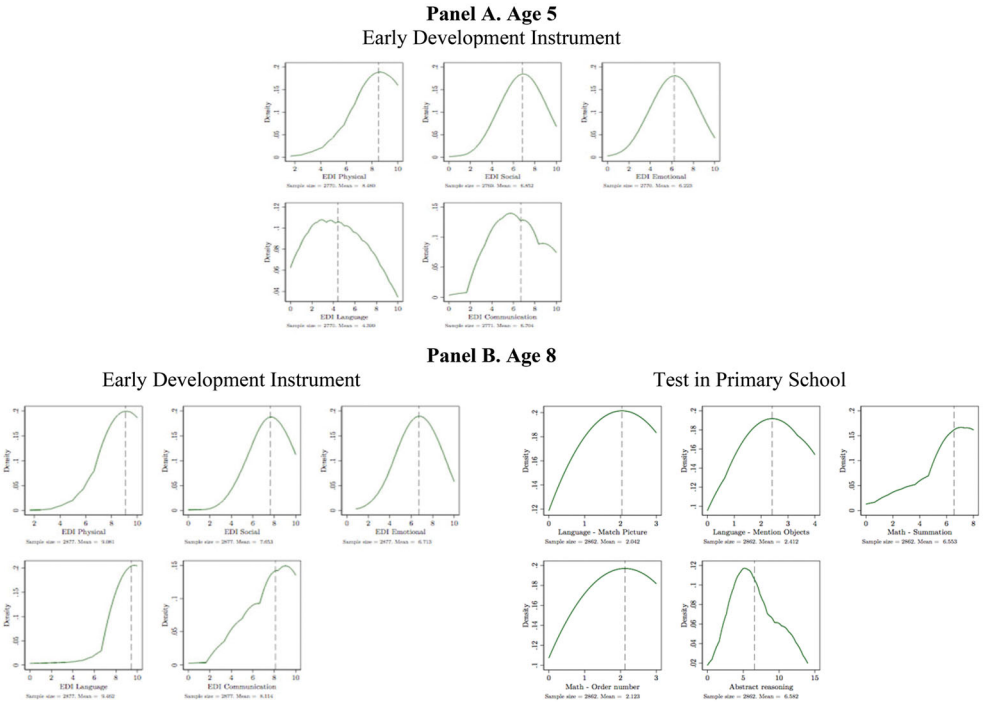

Figure 2. Density of child development outcome measures. Note: Panel A shows the density of the EDI at age 5. Panel B shows the density of the EDI and test scores at age 8. The EDI has a scale of 0 (lowest) to 10 (highest) on five domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication and general knowledge. The test includes five sections: match picture (3 items), mention objects (4 items), summation (8 items), order numbers (3 items), and abstract reasoning (15 items). For both the EDI and the test, higher values indicate better developmental outcomes. Dashed vertical line indicates the mean score.

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.

Figure 2 Panel B shows the distribution of the test scores. In the figure, the x-axis is the number of test items that construct the section of the test. Like the EDI at age 8, the test scores at age 8 exhibit some ceiling effects and may contribute to measurement error. In particular, the math summation section shows a highly skewed distribution with the mean student correctly answering more than 6 items out of 8. 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 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.[14]

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.

For all instruments described above a standard protocol was followed (Pradhan et al., 2013). Questionnaires were first developed in English. This included a translation of all items into Bahasa Indonesia. Questionnaires were then back-translated by a different person to check for accuracy. Any discrepancies were discussed among the research team and addressed. Discussions were also held between the researchers and government counterparts to discuss the relevance and validity of the content of the instruments. Finally, after being trained on the questionnaires, enumerators read and discussed the translated version in their respective local language(s) and made additional changes as a group. All of these changes were recorded in the enumerator manual.

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

[14]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 et al. (2013).

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:

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

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

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/m^2 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. This is to be expected given that the villages were randomly assigned their batch status. As a result, we examine batch 1 and batch 2 villages collectively as treatment villages in our regression specification in Equations (1) and (2).

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

	Treatment				Differences	
	Early (Batch 1) (1)	Late (Batch 2) (2)	Both (Batch 1 and 2) (3)	Comparison (4)	Early – Late (5)	Both – Comparison (6)
Age (years)	1.520	1.499	1.509	1.508	–0.021	0.001
	(0.287)	(0.286)	(0.286)	(0.286)	(0.012)	(0.013)
Household size	4.678	4.729	4.704	4.964	0.052	–0.259*
	(1.529)	(1.568)	(1.549)	(1.705)	(0.095)	(0.102)
Wealth *z*-score (S.D.)	–0.0125	–0.0301	–0.0217	0.0520	–0.018	–0.074
	(1.036)	(0.962)	(0.998)	(1.003)	(0.082)	(0.073)
Parenting *z*-score (S.D.)	–0.0120	–0.0316	–0.0222	0.0532	–0.020	–0.075
	(1.006)	(0.968)	(0.986)	(1.031)	(0.070)	(0.073)
Mother's education is primary or less (1 = Yes)	0.512	0.514	0.513	0.504	0.002	0.008
	(0.500)	(0.500)	(0.500)	(0.500)	(0.029)	(0.030)
Male (1 = Yes)	0.491	0.516	0.504	0.525	0.025	–0.021
	(0.500)	(0.500)	(0.500)	(0.500)	(0.023)	(0.019)
Body Mass Index (BMI)	14.62	14.53	14.57	14.66	–0.092	–0.091
	(2.078)	(2.117)	(2.098)	(2.097)	(0.107)	(0.102)
Cognitive skills (S.D.)	–0.0979	0.00907	–0.0421	0.101	0.107	–0.143*
	(1.055)	(0.982)	(1.019)	(0.947)	(0.076)	(0.061)
Fine motor skills (S.D.)	–0.0323	–0.00911	–0.0202	0.0483	0.023	–0.069
	(0.988)	(1.045)	(1.018)	(0.954)	(0.059)	(0.055)
Gross motor skills (S.D.)	–0.0798	–0.0134	–0.0452	0.108	0.066	–0.153**
	(1.013)	(1.017)	(1.015)	(0.954)	(0.057)	(0.047)
Language skills (S.D.)	–0.0216	–0.00990	–0.0155	0.0371	0.012	–0.053
	(0.984)	(1.031)	(1.008)	(0.979)	(0.058)	(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).

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.

Results

The intent-to-treat impact estimates for the younger cohort are presented in Tables 2–4.[15] 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.[16] When interpreting these results, it is important to note that the counterfactual is

[15]Results by batch are reported in Appendix Tables 1–3.

[16]In both cases—wealth and parenting practices—we split the sample into those above the mean and those below the mean.

Table 2. Impact on enrollment outcomes for the younger cohort.

Outcome	Survey Year		All (1)	Poor (2)	Non-poor (3)	Low parenting score (4)	High parenting score (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
		Obs.	2,778	1,201	1,577	1,540	1,238
	2016	Coeff.	0.498***	0.539***	0.470***	0.466***	0.532***
		(S.E.)	(0.023)	(0.029)	(0.027)	(0.027)	(0.026)
		Comp. mean	0.0673	0.0499	0.0802	0.0686	0.0659
		Obs.	2,894	1,289	1,605	1,834	1,060
Ever enrolled in non-project playgroup until survey year	2013	Coeff.	−0.220***	−0.193***	−0.241***	−0.215***	−0.223***
		(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
	2016	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
		Obs.	2,894	1,289	1,605	1,834	1,060
Ever enrolled in kindergarten until survey year	2013	Coeff.	−0.075	−0.074	−0.085*	−0.090	−0.059
		(S.E.)	(0.040)	(0.053)	(0.041)	(0.048)	(0.043)
		Comp. mean	0.532	0.441	0.598	0.522	0.542
		Obs.	2,778	1,201	1,577	1,540	1,238
	2016	Coeff.	−0.086*	−0.068	−0.098*	−0.098*	−0.075
		(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 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
		Obs.	2,778	1,201	1,577	1,540	1,238
	2016	Coeff.	8.683***	8.840***	8.599***	8.365***	9.022***
		(S.E.)	(0.413)	(0.544)	(0.469)	(0.496)	(0.491)

(*continued*)

Table 2. Continued.

Outcome	Survey Year		All (1)	Poor (2)	Non-poor (3)	Low parenting score (4)	High parenting score (5)
		Comp. mean	0.327	0.321	0.331	0.320	0.334
		Obs.	2,894	1,289	1,605	1,834	1,060
Months enrolled in non-project playgroup until survey year	2013	Coeff.	−3.001***	−2.094***	−3.743***	−2.721***	−3.284***
		(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
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
		Obs.	2,778	1,201	1,577	1,540	1,238
	2016	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.

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.

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.

children living in comparison villages, who may or may not have access to other services such as non-project playgroups or kindergartens.

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.[17] 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% 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, suggesting that parents to some extent view playgroups as substitutes for kindergartens.[18]

The results for months of enrollment are largely consistent with our findings for enrollment rates. The project increased children's enrollment duration in project

[17]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.

[18]Our experience in the field and the data on enrollment histories does indeed underscore the fact that parents view playgroups and kindergartens as substitutes and do not adhere to the ages of eligibility for the different services.

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.[19] 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.[20] 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 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.

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). Our results suggest that the impact on emotional maturity may be concentrated among those classified as poor in our data.[21] 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. As noted earlier, measurement error may contribute to the null effects on the EDI that we observe in 2016. In the raw densities of the EDI domains presented in Figure 2, 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.

[19]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.

[20]See Appendix Table 4.

[21]See Appendix Table 5.

In Table 4, we find mixed results of the project on primary school test scores, which were measured at age 8 (2016). We find moderate positive intent-to-treat 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 moderate positive intent-to-treat 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.[22] 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.[23]

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., 2017). In this section, we contrast the impact estimates between the younger and older cohort, which varied in terms of (1) the phase of project implementation when they were exposed to playgroups and (2) the length of exposure to playgroups at an *appropriate* age. The novel data available to us allows us to paint a more complete picture of the relationship between these factors and child development in the longer-term.

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.[24] 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 2–4 in this paper. Columns 2 and 5 present the equivalent results for the older cohort, previously reported in Brinkman, Hasan, Jung, Kinnell and Pradhan (2017). Column 6 presents test score results for the older cohort, which have not previously been published.[25] Columns 3 and 7 show the results of a *t*-test comparing the differences in impact between the two cohorts.[26] We plot these columns in Figure 3.

Overall, we find that the project "impact" differed between the older and younger cohorts. By age 5, children from the younger cohort were 34.6 percentage points more

[22]Appendix Table 6.
[23]See Appendix Figure 2.
[24]See Appendix B for the empirical strategy used to estimate the impacts for the 4 year-old cohort.
[25]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.
[26]Our test statistic to examine whether the estimates from the two cohorts are statistically different is $\frac{\widehat{\beta_{1yr}}-\widehat{\beta_{4yr}}}{SE(\widehat{\beta_{1yr}}-\widehat{\beta_{4yrs}})} = \frac{\widehat{\beta_{1yr}}-\widehat{\beta_{4yr}}}{\sqrt{Var\left(\widehat{\beta_{1yr}}\right)+Var\left(\widehat{\beta_{4yr}}\right)-2Cov(\widehat{\beta_{1yr}},\widehat{\beta_{4yr}})}} = \frac{\widehat{\beta_{1yr}}-\widehat{\beta_{4yr}}}{\sqrt{SE\left(\widehat{\beta_{1yr}}\right)^2+SE\left(\widehat{\beta_{4yr}}\right)^2}}$ assuming $2Cov\left(\widehat{\beta_{1yr}},\widehat{\beta_{4yr}}\right)=0$. Since $Cov\left(\widehat{\beta_{1yr}},\widehat{\beta_{4yr}}\right)$ is typically positive, our assumption yields a conservative estimate of the test statistic.

likely to have ever been enrolled in project playgroups than children from the older cohort. By age 5, they were 17.6 percentage points less likely to have ever enrolled in non-project playgroups and there was no practical or statistical difference between the two cohorts in ever being enrolled in kindergarten by age 5. This pattern is also visible at age 8 though the magnitudes are smaller. The younger cohort was 22.9 percentage points more likely to have ever enrolled in project playgroups than the older cohort and 15.2 percentage points less likely to have enrolled in non-project playgroups. Even by age 8, there is not discernible difference in kindergarten enrollment rates between the two cohorts.

In terms of months of enrollment the pattern is the same. The younger cohort was enrolled in project playgroups for longer—both by age 5 (6 months) and by age 8 (5 months) than the older cohort. Similarly the younger cohort was enrolled in non-project playgroups for a shorter duration—both by age 5 and by age 8. There is no statistical difference in their duration of enrollment in kindergartens by either age.

Our ability to draw conclusions from the Early Development Instrument results is constrained by the fact that the instrument starts to display ceiling effects as early as age 5 in some domains (refer to Figure 2). Thus, even though we see some statistically significant differences between the two cohorts in physical health and well-being and social competence by age 5, we refrain from interpreting these as indicative of a difference in the experience of the two cohorts. By age 8 no differences remain.

However, the test score data provide some insights into whether greater exposure to the project by the younger cohort results in improved child development outcomes. There is evidence that the younger cohort did substantially better than the older cohort—0.23 SD better in language when matching pictures to word, 0.28 SD better in mathematics when ordering numbers. While the point estimate on summation is positive and significant, this measure also displays some ceiling effects. Given that these tests were administered in the early grades of primary school, we also take these results as indication of the fact that the impacts of this intervention lasted beyond the duration of the project.

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 ages (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.[27]

[27]As a robustness check, we also examine whether treatment effects on the EDI and test scores vary by exposure to playgroups at the appropriate ages among the younger cohort. Results are reported in Appendix Tables 5 and 6. Overall, we find larger treatment effects for social competence and emotional maturity among children enrolled in playgroups for one to two years at ages 3–4 relative to their peers enrolled in playgroup for less than one year at ages 3–4. These results are consistent with the argument we present here that longer exposure to playgroup at the appropriate age explains part of the different impacts we observe between the younger and older cohorts.

Table 5. Comparison of impact estimates for the two cohorts.

	Treatment effect at age 5			Treatment effect at age 8			
	Younger cohort (1)	Older cohort (JOLE results) (2)	Difference between cohorts (3)	Younger cohort (4)	Older cohort (JOLE results) (5)	Older cohort (new results) (6)	Difference between cohorts (7)
Ever enrolled in project playgroup	0.499***	0.153***	0.346***	0.498***	0.269***		0.229***
	(0.022)	(0.015)	(0.027)	(0.023)	(0.018)		(0.029)
Ever enrolled in non-project playgroup	−0.220***	−0.044**	−0.176***	−0.221***	−0.069***		−0.152***
	(0.037)	(0.014)	(0.040)	(0.037)	(0.017)		(0.041)
Ever enrolled in kindergarten	−0.075	−0.058*	−0.017	−0.086*	−0.130***		0.044
	(0.040)	(0.026)	(0.048)	(0.040)	(0.028)		(0.049)
Months in project playgroup	7.781***	1.914***	5.867***	8.704***	3.855***		4.849***
	(0.396)	(0.158)	(0.426)	(0.414)	(0.254)		(0.486)
Months in non-project playgroup	−3.001***	−0.647***	−2.354***	−3.274***	−0.943***		−2.331***
	(0.637)	(0.156)	(0.656)	(0.645)	(0.238)		(0.688)
Months enrolled in kindergarten	−0.199	0.590*	−0.789	−1.037	−1.546***		0.509
	(0.436)	(0.270)	(0.513)	(0.664)	(0.444)		(0.799)
Physical health and well-being (SD)	0.208***	−0.026	0.234*	0.002	0.104		−0.102
	(0.051)	(0.076)	(0.091)	(0.064)	(0.074)		(0.098)
Social competence (SD)	0.018	0.223**	−0.205*	−0.012	0.024		−0.036
	(0.053)	(0.076)	(0.093)	(0.049)	(0.075)		(0.09)
Emotional maturity (SD)	0.115*	0.014	0.101	0.003	0.158*		−0.155
	(0.056)	(0.071)	(0.090)	(0.045)	(0.068)		(0.082)
Language & cognitive development (SD)	0.073	0.128	−0.055	0.035	0.056		−0.021
	(0.056)	(0.070)	(0.090)	(0.040)	(0.060)		(0.072)
Communication & general knowledge (SD)	−0.005	0.075	−0.080	−0.160***	0.014		−0.174
	(0.074)	(0.079)	(0.108)	(0.061)	(0.132)		(0.145)
Language—match picture (SD)				0.133**		−0.100	0.233***
				(0.058)		(0.058)	(0.082)
Language—mention objects (SD)				−0.027		−0.050	0.023
				(0.053)		(0.054)	(0.076)
Math—summation (SD)				0.068		−0.148**	0.216***
				(0.054)		(0.051)	(0.074)
Math—order numbers (SD)				0.125**		−0.155*	0.280***
				(0.058)		(0.061)	(0.084)
Abstract reasoning (SD)				−0.021		−0.124*	0.103
				(0.044)		(0.057)	(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.

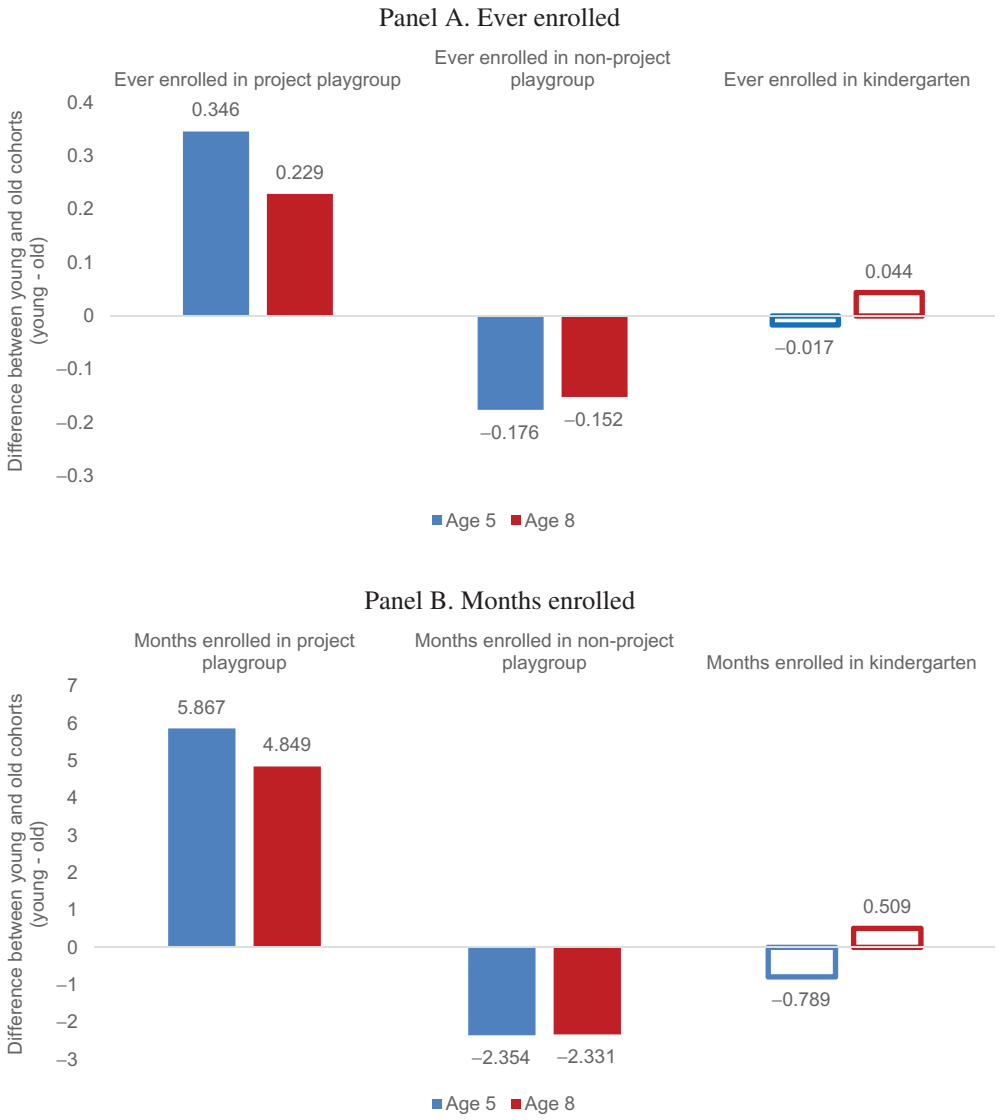

Figure 3. Differences in impacts between the younger cohort and older cohort. Notes: Figures show results in columns 3 and 7 in Table 5. For Panels A-C, results are shown at age 5 (blue) and age 8 (red). For Panel D, results are only available at age 8. Solid bars denote statistically significant differences at the 5% level. Hollow bars are not statistically significant. Labels in bold denote significance at the 1% level while those in italics denote significance at the 5% level.

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, costs approximately US$27 per child (all amounts in 2014 dollars).

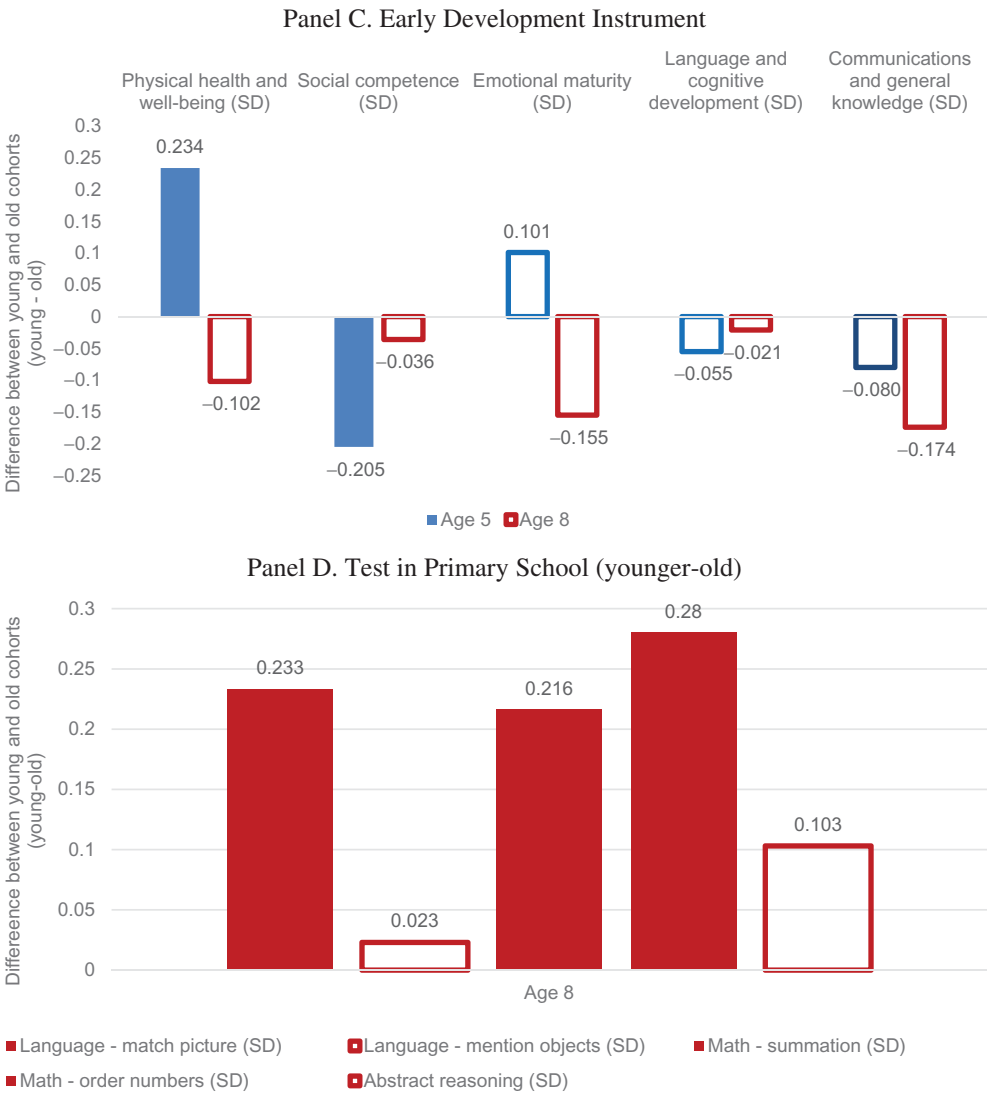

Figure 3. Continued

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 to 12.3% as estimated by Patrinos et al. (2006). We assume that:

a. there is a 6.5% 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)

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.

b. children do not begin to realize the benefits of increased wages until age 18
c. they do so for 40 years

Under conservative assumptions, a 0.1-year increase in schooling results in a benefit-cost ratio of 0.65.[28] 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 both published literature and the World Bank Project Appraisal Document (11.2%) suggests a correspondingly much higher benefit-cost ratio of 1.12 − 7.84. Thus reasonable 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.

Discussion and Conclusion

This paper examined treatment effect variation across two cohorts of children who were exposed to a project that expanded access to playgroups at different points in the project's lifespan. Although the project had positive impacts for both cohorts of children, these impacts varied across cohorts. This suggests that the maturity and timing of pre-primary education services are important factors to consider not only when designing and implementing impact evaluations but also when designing policy.

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.

[28]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.

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).[29] 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 higher wealth z-scores than those in the older cohort.[30] 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.[31]

Likewise, the quality of the playgroups likely ebbed and flowed during the period under study. Brinkman et al. (2017) 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., 2019). 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, totally 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 Figure 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.

[29]See Appendix Table 7. The median monthly fee was 5,000 IDR in 2010 and 10,000 IDR in 2013. Rural households at the time reported a monthly wage of 1.7 million IDR (Bureau of Labor Statistics, 2016).
[30]See Appendix Table 8.
[31]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., 2019).

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 results suggest that 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 pre-primary education 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 delivered to successive cohorts of beneficiaries.

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].

References

Attanasio, O., & Vera-Hernández, M. (2004). Medium and long-run effects of nutrition and child care: Evaluation of a community nursery program in rural Colombia (IFS Working Paper EWP04/06). Institute for Fiscal Studies.

Barnett, W. S. (1997). Costs and financing of early child development programs. In Mary E. Young (Ed), *Early child development. Investing in our children's future* (pp. 287–295). Elsevier.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*(6045), 975–978. https://doi.org/10.1126/science.1204534

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, *92*(5–6), 1416–1440.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of American Medical Association*, *288*(6), 728–737. https://doi.org/10.1001/jama.288.6.728

Bouguen, A., Filmer, D., Macours, K., & Naudeau, S. (2018). Preschool and parental response in a second best world: Evidence from a school construction experiment. *Journal of Human Resources*, *53*(2), 474–512.

Brennan, L. M., Shaw, D. S., Dishion, T. J., & Wilson, M. (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), 1289–1300. https://doi.org/10.1007/s10802-012-9639-2

Brinkman, S. A., Kinnell, A., Maika, A., Hasan, A., Jung, H., & Pradhan, M. (2017). Validity and reliability of the early development instrument in Indonesia. *Child Indicators Research*, *10*(2), 331–352. https://doi.org/10.1007/s12187-016-9372-4

Brinkman, S., Hasan, A., Jung, H., Kinnell, A., & Pradhan, M. (2015). *The impact of expanding access to early childhood services in rural Indonesia: evidence from two cohorts of children*. The World Bank.

Brinkman, S., Hasan, A., Jung, H., Kinnell, A., & Pradhan, M. (2017). The impact of expanding access to early childhood education services in rural Indonesia. *Journal of Labor Economics*, *35*(S1), 305–335.

Brinkman, S., Hasan, A., Jung, H., Kinnell, A., Nakajima, N., & Pradhan, M. (2017). The role of preschool quality in promoting child development: evidence from rural Indonesia. *European Early Childhood Education Research Journal*, *25*(4), 483–505. https://doi.org/10.1080/1350293X.2017.1331062

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, *25*(2), 140–165. https://doi.org/10.1016/j.ecresq.2009.11.001

Conti, G., & Heckman, J. (2012). The economics of child well-being (NBER Working Paper. 18466). NBER.

Cortázar Valdés, A. (2011). *Long-term effects of early childhood education on academic achievement in Chile* [Ph.D. dissertation]. Teachers College, Columbia University (ProQuest AAT 3492908).

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, *91*(4), 795–813.

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives : a Journal of the American Economic Association*, *27*(2), 109–132. https://doi.org/10.1257/jep.27.2.109

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., de Mello, M. C., Hidrobo, M., Ulkuer, N., Ertem, I., & Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339–1353. https://doi.org/10.1016/S0140-6736(11)60889-1

Evans, J. L., Myers, R. G., & Ilfeld, E. M. (2000). *Early childhood counts: A programming guide on early childhood care for development*. The World Bank.

Feinstein, L., & Duckworth, K. (2006). *Development in the early years: Its importance for school performance and adult outcomes*. Centre for Research on the Wider Benefits of Learning Institute of Education.

Hasan, A., Hyson, M., & Chang, M. C. (Ed.). (2013). *Early childhood education and development in poor villages of Indonesia: Strong foundations, later success*. The World Bank.

Hasan, A., Jung, H., Kinnell, A., Maika, A., Nakajima, N., & Pradhan, M. P. (2019). Built to last: Sustainability of early childhood education services in rural Indonesia (World Bank Policy Research Working Paper No. 9061). World Bank.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

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. https://doi.org/10.1146/annurev-publhealth-031912-114500

Janus, M., & 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, H., & Hasan, A. (2016). The impact of early childhood education on early achievement gaps in Indonesia. *Journal of Development Effectiveness*, *8*(2), 216–233. https://doi.org/10.1080/19439342.2015.1088054

Martinez, S., Naudeau, S., & Pereira, V. (2017). *Preschool and child development under extreme poverty: Evidence from a randomized experiment in rural Mozambique*. The World Bank.

Melhuish, E. C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M. B., & Malin, A. (2008). The early years. Preschool influences on mathematics achievement. *Science*, *321*(5893), 1161–1162. https://doi.org/10.1126/science.1158808

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Moser, S. E., West, S. G., & Hughes, J. N. (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), 603–621. https://doi.org/10.1037/a0027571

Nakajima, N., Hasan, A., Jung, H., Brinkman, S., Pradhan, M., & Kinnell, A. (2019). Investing in school readiness: A comparison of different early childhood education pathways in rural Indonesia. *International Journal of Educational Development*, *69*, 22–38. https://doi.org/10.1016/j.ijedudev.2019.05.009

Nores, M., & Barnett, S. (2010). Benefits of early childhood interventions across the world: (under) Investing in the very young. *Economics of Education Review*, *29*(2), 271–282. https://doi.org/10.1016/j.econedurev.2009.09.001

Office of Population Studies, University of SanCarlos. (2005). *A study of the effects of early childhood interventions on children's physiological, cognitive and social development.*

Patrinos, H. A., Ridao-Cano, C., & Sakellariou, C. (2006). *Estimating the returns to education: Accounting for heterogeneity in ability*. The World Bank.

Pradhan, M., Brinkman, S. A., Beatty, A., Maika, A., Satriawan, E., de Ree, J., & Hasan, A. (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*(1), 259 https://doi.org/10.1186/1745-6215-14-259

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academy Press.

Shonkoff, J. P. (2014). Changing the narrative for early childhood investment. *JAMA Pediatrics*, *168*(2), 105–106. https://doi.org/10.1001/jamapediatrics.2013.4212

Stevens, K. B., & English, E. (2016). Does pre-K work? The research on ten early childhood programs – And what it tells us. *AEI Paper & Studies*, COVd.

Tanner, J. C., Candland, T., & Odden, W. S. (2015). *Later impacts of early childhood interventions: A systematic review*. The World Bank.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., Sachdev, H. S. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *The Lancet*, *371*(9609), 340–357. https://doi.org/10.1016/S0140-6736(07)61692-4

Wasik, B. A., & Snell, E. K. (2019). Synthesis of preschool dosage: How quantity, quality, and content impact child outcomes. In A. J. Reynolds & J. A. Temple (Eds.), *Sustaining early*

childhood learning gains: Program, school, and family influences (pp. 31–51). Cambridge University Press.

Wolf, S., Aber, J. L., Behrman, J. R., & Tsinigo, E. (2019). Experimental impacts of the "Quality preschool for Ghana" interventions on teacher professional well-being, classroom quality, and children's school readiness. *Journal of Research on Educational Effectiveness*, *12*(1), 10–37. https://doi.org/10.1080/19345747.2018.1517199

Zubrick, S. R., Smith, G. J., Nicholson, J., Sanson, A., & Jackiewicz, T. A. (2008). Parenting and families in Australia. *FaHCSIA Social Policy Research Paper*.