

Improving Preschool Provision and Encouraging Demand

Evidence from a Large-Scale Construction Program

Jan Berkes
Adrien Bouguen
Deon Filmer
Tsuyoshi Fukao



WORLD BANK GROUP

Development Economics
Development Research Group

&

Education Global Practice

December 2019

Abstract

This paper studies the impact of a preschool construction program and of two demand-side interventions in Cambodia. Within this context where other preschools are available, impacts are likely to differ between children who would have been enrolled in a preexisting preschool and those who would have stayed at home, with larger expected gains for the latter. The construction program caused enrollment to increase but demand-side interventions did not. After one year, the study measures intent-to-treat impacts on cognitive (0.04 standard deviations) and socio-emotional development (0.07 SD). The analysis also shows that the effect on children who would have stayed at home can be

bounded (between 0.14 SD and 0.45 SD). With further assumptions these bounds can be tightened (0.14 SD – 0.35 SD) and under heavier assumptions, the study estimates this effect to be 0.19 SD, while the effect on children who would have benefited from another preschool is small and insignificant. These results are consistent with measures of preschool quality which imply that the newly constructed schooled only significantly improved infrastructure quality and not process quality. After two years of program implementation, most impacts become insignificant suggesting that the advantage provided by preschool quickly vanished, specifically once children enrolled in primary school.

This paper is a product of the Development Research Group, Development Economics and 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 dfilmer@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.

Improving Preschool Provision and Encouraging-Demand: Evidence from a Large-Scale Construction Program*

Jan Berkes[†], Adrien Bouguen[‡], Deon Filmer[§], Tsuyoshi Fukao[§]

Originally published in the [Policy Research Working Paper Series](#) on *December 2019*.
This version is updated on *February 2022*.
To obtain the originally published version, please email prwp@worldbank.org.

Keywords: early childhood development; preschool; cognitive skills; socio-emotional skills

JEL Codes: I2,I3

*This research project is a joint effort between the Cambodian government, the World Bank, a team of field researchers (Angkor Research), and a team of academic researchers. We are particularly grateful to the World Bank for its constant support: special thanks to Simeth Beng. This research was funded by SIEF. Special thanks to Alaka Holla for her useful comments throughout the program's duration. The project could not have been conducted without the collaboration of the Ministry of Education, Youth and Sport in Cambodia. We are particularly grateful to Sok Sokhom and Lynn Dudley. Data were collected in the field by a team of very dedicated field researchers from Angkor Research: special thanks to John Nicewinter, Ian Ramage, Benjamin Lamberet, Kimhorth Keo, and Ratanaksophea Saing. Finally, many researchers also contributed to this research through their very useful comments: Craig McIntosh, Patrick Kline, Christopher Walters, Karen Macours, Diego Vera, Markus Frölich, Paul Gertler, Supreet Kaur, Edward Miguel, Katja Kaufmann, Elisabeth Sadoulet, Analia Schlosser, Alain de Janvry, Clément de Chaisemartin, Antoine Camous, Harald Fadinger, Luc Behaghel, Clement Imbert, and Marc Gurgand. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent. The study was preregistered at AEA's Social Science Registry (AEARCTR-0001045).

[†]DIW Berlin, Federal Ministry of Labour and Social Affairs Germany

[‡]Santa Clara University. Adrien Bouguen also acknowledges financial support from the German Research Foundation (DFG) project SFB 884 during his stay at the University of Mannheim

[§]The World Bank

1 Introduction

Robust early childhood development lays the foundation for greater human capital development in later childhood and beyond (Heckman, 2006; Cunha and Heckman, 2007; Almond et al., 2018; World Bank, 2018). Enrollment rates in pre-primary education are growing worldwide (UNICEF, 2019)—a momentum supported by strong evidence that early child nurture and stimulation can lead to positive impacts on later outcomes (Shonkoff et al., 2016; Black et al., 2017; Britto et al., 2017). While preschool programs can be highly effective, much of this evidence is based on relatively small targeted programs (Gertler et al., 2014; Elango et al., 2015). Whether positive effects on early development can be sustainably engendered through large-scale programs is less clear, particularly when the quality of services is low (Engle et al., 2011; Britto et al., 2011; Ichino et al., 2019; Andrew et al., 2019) or services are not accessed by children who might benefit from them (Cornelissen et al., 2018). The issue of scalability concerns high-income as well as low- and middle-income countries (List et al., 2021).

In this study, we evaluate the impacts of a large-scale preschool construction program on school participation and child development in a low-income country setting, Cambodia, using a randomized controlled trial. The main objective of the construction program was to provide a quality preschool experience to every child in the treatment villages who met the age criteria (3-5 years old). We assess whether the program and supplementary interventions increased enrollment, parental involvement and, ultimately, child cognitive and socio-emotional development. Our evaluation has three treatment arms: the construction of formal community preschools (CPS) (which we refer to as T1); the addition of a demand-side intervention to promote awareness about CPS and the value of education (T2); and the further addition of a home-based program consisting of trained “core parents”, who provide caregiver training sessions focused on good parenting, the value of nutrition, and the importance of preschool (T3). The 305 study villages were randomly assigned either to a control group (C) or one of the treatment arms.

The CPS program was implemented in a context where alternative preschools were available, in particular informal preschool programs. In most cases, the CPS caused the informal programs to shut down. The CPS program therefore implicitly generated two types of compliers: the children who would have attended an alternative program absent the construction (*a-compliers*) and children who would have stayed at home (*h-compliers*), with impacts larger for the latter. Using traditional treatment parameters (ITT or LATE), the effect on *a-compliers* (later referred to as $LATE_{ac}$) and the effect on *h-compliers* ($LATE_{hc}$) are not identified but are crucial for policy. Indeed, the effect on both complier types (also referred to as subLATEs) corresponds to the two main objectives of the CPS program. The $LATE_{hc}$ captures the effect on the children who would not have had any access to school otherwise,

while the $LATE_{ac}$ measures how well the new CPS program fares in comparison with the existing offer. Measuring the impacts on both subLATEs is therefore crucial to understanding how the program affects outcomes, and therefore for proving policy recommendations.

We study the reduced form effects for each treatment group using midline (1 year after construction) and endline (2 years after construction) data, but also the effects on both sub-groups of compliers. The reduced form effects are straightforward to derive, empirically valid, and policy relevant as they reflect the program’s impacts in a typical Cambodian preschool environment. The subLATE parameters are not identified but the $LATE_{hc}$ can be bounded. In addition, under heavier assumptions, the subLATEs can be estimated using additional instruments interacted with the treatment variable.

We have four main findings. First, the construction of CPS—the implementation of which closely followed the program and experimental design—increased preschool participation. Children in villages with a newly built CPS were about 52 percentage points (pp) more likely to have ever attended a CPS by the time they were between four and six years old (or 11 pp to have ever attended any preschool). Put differently, they were enrolled, on average, about 4.5 months longer than children in the control group.

Second, we find significant positive ITT impacts one year after the program started but these effects disappear after two years. More specifically, children in treatment villages scored about 0.04 SD higher on an index of cognitive development and about 0.07 SD higher on an index of socio-emotional development, one year after the program’s implementation. These ITT impacts are consistent with moderate-to-large impacts on children who would have stayed at home absent the program ($LATE_{hc}$). We show that this subLATE parameter can be bounded between 0.14 SD and 0.46 SD on our cognitive development measures using light assumptions; with additional assumptions the bounds can be narrowed to between 0.14 and 0.35 SD and, under even more restrictive assumptions, the $LATE_{hc}$ can be estimated to be around 0.19 SD. We also estimate the effect on children who would have attended another preschool absent the program ($LATE_{ac}$) to be consistently close to zero, confirming that the quality of education provided in the newly built preschools was not fundamentally improved. These results are consistent with direct measures of school and teacher quality that indicate that the CPS construction program improved preschool quality mostly in dimensions that tend to be weakly correlated with child performance (e.g. equipment, materials, quality of the building—which we refer to as “structural quality”) while it did not improve the quality of educational processes (e.g. teacher quality or teaching practices—which we refer to as “process quality”). After two years (at endline), however, these impacts were no longer statistically significant for the full sample. As discussed below, the

lack of persistence of early impacts is consistent with results found in other contexts and is not necessarily inconsistent with potential long-lasting impacts.

Third, these overall results may mask some variation. While preschool is sometimes suspected to be a substitute for parental involvement in high income countries (Baker et al., 2008), in our context, the impacts on cognitive development for children from the wealthiest or most educated quartile of households were slightly larger and sometimes still measurable at endline, suggesting that for relatively more privileged students the impact of the program is not only larger but is also longer-lasting. Our statistical precision is however low in sub-group analysis: our results are only barely significant and would not pass a multi-hypothesis test. Yet, if taken at face value, our results suggest that less disadvantaged families were better able to take advantage of the CPS, through higher level of preschool attendance, higher level of parental involvement and possibly a home environment that complemented formalized education.

Fourth, we find that enrollment rates were not different across the treatment arms. Children in villages where the demand-side program (which we refer to as door-to-door, or D2D) and home-based-program (HBP) interventions were deployed were not systematically more likely to enroll in CPS or perform better in terms of developmental measures. This was despite the fact that all respondents in T2 and T3 received the information leaflet (as it was part of the baseline and midline surveys) and almost all T3 villages benefited from a home-based program (95%). Yet, in T2 and T3 villages, caregivers rarely remember receiving the leaflet (+ 8 pp compared to the control), did not recall receiving more door-to-door visits than the control group and did not report much more participation in home-based program sessions (+10 pp). Our findings therefore suggest that the information and HBP interventions did not modify the effects of CPS: they did not significantly increase enrollment and did not modify the performance of the CPS. Parents did not update their reported estimate of the return to education either. Interestingly, our results show that parents were sufficiently informed about the CPS and did not underestimate the return to education as shown in other contexts (Nguyen, 2008; Jensen, 2010). Other reasons (e.g. distance, short CPS shift duration) seem to constitute more binding constraints.

There are various potential reasons why we find relatively modest short-term CPS impacts which fade out after two years for most children. First, while our sub-LATE results suggest positive $LATE_{hc}$, $LATE_{ac}$ are close to zero. This suggest that the CPS did not fundamentally improve the quality of the (existing) preschool education provided. Relatedly, we find evidence that the significant improvements in structural quality was not accompanied by a concomitant improvement in process quality. Classroom observations show that CPS were substantially better than preexisting preschools in terms of the availability of infrastructure and materials, but they also

show that curriculum content and quality of pedagogy, as well as the frequency and quality of teacher-child interactions, were not significantly better in CPS than in the previous informal preschools.¹

Second, the absence of effect two years after the CPS construction is in part explained by complex enrollment patterns. Indeed, the total number of months of exposure to (any) school is as high at midline as at endline. This surprising result is explained by the large proportion of children (treatment and control) already enrolled in primary school at endline (therefore not contributing to the overall enrollment effect) while the additional preschool enrollees are in their vast majority *a-compliers*, i.e. they would have enrolled in an alternative preschool anyway and therefore do not increase overall treatment effect on enrollment either. Since our subLATE analysis reveals that the $LATE_{ac}$ is close to zero, it is not surprising that the overall endline treatment effect would be driven towards zero. The only children who could have driven up the effect at endline are the children who had enrolled in a CPS at midline instead of home care, and are enrolled in a primary school at endline. The lack of average endline impacts therefore indicates that the small benefits from attending CPS the first year do not translate into longer-term advantages in primary school. Once everyone is enrolled in primary school, the children who did not benefit from a preschool appear to catch up fairly quickly.²

This finding does not necessarily rule out potential longer-term benefits of having started school earlier. Evidence from the United States suggests that medium-term fade-out might be consistent with long-term improved outcomes (Heckman et al., 2013) perhaps because other—non-measured—aspects of child development improved in the short-run (Duncan and Magnuson, 2013). In addition, other research suggests that reaping the full medium- and long-term benefits of preschool might require complementary investments at later stages of human capital development. For example, it is possible that high-quality primary schools, which may be in short supply in Cambodia, would be required in order to sustain impacts.³ If wealthier or more educated households are more able to invest in complementary and high-quality inputs during preschool and then into the primary school years, this might explain the sustained impacts on cognitive development for children from those households. Further research that tracks individuals over longer periods of time would be necessary to investigate these various hypotheses.

Our paper contributes to a limited but growing literature on preschool availability

¹These results are consistent with the literature showing that pedagogical practices and the quality of teacher-child interactions are particularly important for child development (Araujo et al., 2016, 2019; Andrew et al., 2019).

²There is an additional reason which we return to: the preschool infrastructure availability at baseline was slightly lower in the treatment groups than in the control group. The fact that this difference was significant for T2 explains in large part why the treatment T2 shows smaller impacts than the other treatment branches.

³Johnson and Jackson (2018) document such dynamic complementarities between the Head Start program and investments in primary and secondary schools in the United States.

that finds mixed results. While positive short-term effects on child development are reported for Mozambique (Martinez et al., 2017b), other studies find insignificant or even negative effects (Brinkman et al., 2017; Bouguen et al., 2018; Blimpo et al., 2019; Bernal et al., 2019).⁴ Differences in socio-economic conditions, along with the nature and quality of a preschool program and its counterfactual (quality of other preschool arrangements or quality of parenting), make it difficult to draw broad conclusions that are applicable across contexts, particularly when these differences are unobservable or undocumented. Specifically, the literature often lacks detailed information on both the quality of the preschool supply and a clear understanding of the preschool demand mechanisms. Similar issues exist for the literature in higher income countries where large and long-lasting impacts (Belfield et al., 2006; Carneiro and Ginja, 2014) coexists with disappointing (Puma et al., 2012) and even negative results (Baker et al., 2008, 2019).

The presence of alternative programs may, in particular, explain some of the lack of consistency in the literature on preschool impacts. Most large-scale randomized controlled trials aimed at measuring preschool effects in low-income countries are implemented in an environment where (non-household-based) alternative care arrangements are present. For instance, in a previous preschool experiment conducted in Cambodia from 2008 to 2010, Bouguen et al. (2018) find that 11% of the control group attended a preschool. Similarly, 8% of a control group in Mozambique attended preschool (Martinez et al., 2017a), 16% did so in The Gambia (Blimpo and Pugatch, 2017), and 15% in Indonesia (Brinkman et al., 2017). In the US, 40% of families that lost a lottery to enroll in Head Start ultimately benefited from a close substitute program (Puma et al., 2012). The fact that all of these studies implicitly have different degrees of substitution, along with the fact that the quality of alternative childcare programs is often unknown, makes it impossible to draw general conclusions regarding the effectiveness of preschool interventions. We suspect that this lack of consistency in the literature is at least partially a result of variability in the type of specific substitution patterns we document in this study.

We contribute to the literature in four different dimensions. First, we provide evidence on a large, well-implemented public preschool construction program in rural Cambodia, with detailed information on both children and on the quality of the new preschool services provided. The key outcomes of interest are preschool and subsequent school enrollment, along with child cognitive and socio-emotional development. We collected three rounds of data on about 7000 children to analyze impacts.

Second, using T2 and T3 treatment branches, we analyze the effect of strategies to increase preschool demand and parental involvement. In particular, preschool

⁴Evidence on preschool programs in low- and middle-income countries based on non- and quasi-experimental study designs generally points to positive effects of preschool attendance (Berlinski et al., 2008, 2009; Engle et al., 2011; Rao et al., 2014).

construction was complemented with two additional interventions: (i) a pure demand-side intervention that included a door-to-door program and the provision of information about the preschool and its potential returns (D2D) and a home-based program (HBP), that aimed to increase preschool enrollment by increasing caregivers' awareness about the availability and importance of preschool education as well as increase caregivers' involvement in their children's education.

Third, using detailed classroom observations, we document the quality of the new preschools in comparison to the available alternatives. While most of the experimental or quasi-experimental studies on preschool programs provide some indicators of structural quality, process quality is observed less often. This is a limitation since process quality seems particularly predictive of child development (Araujo et al., 2016, 2019). In a large experiment in Colombian public preschools, Andrew et al. (2019) show that pedagogical training improved the learning environment and children's cognitive development while improvements in structural quality alone did not. Through detailed classroom observations, we are able to explore the degree to which the establishment of the new preschools improved structural and process quality of available preschools in Cambodia and how this mediates the effectiveness of the intervention.

Fourth, using additional empirical strategies and detailed information about the alternative forms of preschool available to parents, we isolate the impact of the program on children who switch to preschool from home care from the impact on those who switch from alternative preschools. We argue that the impact on the former is a critical parameter in the early childhood development program evaluation literature and that failure to isolate both sub-treatment effect parameters contributes to the ongoing confusion in the debate about the effectiveness of preschool availability in low-income countries.

The paper is structured as follows. Section 2 provides some context and background, and then details of the program; Section 3 describes our evaluation design and the data we use in our analysis (including information on covariate balance across treatment arms and on attrition); Sections 4 and 5 respectively outline our empirical strategy and discuss the findings. A final brief section concludes.

2 Background and Program Design

2.1 Early Childhood Development programs in Cambodia

Despite two decades of robust economic growth, Cambodia remains one of the least developed countries in Southeast Asia, with a GDP per capita estimated at \$1,160 in 2015 (roughly \$3,000 in PPP terms). The country also faces multiple challenges in the education sector. With a preschool enrollment rate in 2009 of 40% among five-year-olds (MoEYS, 2014), Cambodia fares poorly in comparison to neighboring

Thailand and Vietnam.⁵ To increase the capacities and quality of its education system, the Cambodian government received a first grant from the Global Partnership for Education (GPE I) of \$57 million for the period 2008—2012, which included an expansion of the early education system (formal preschools, informal preschools, and parenting programs). Bouguen et al. (2013) evaluated the impact of the early childhood education and development (ECD) components of GPE I on child development outcomes. They find no impacts on outcomes overall, although the study period was marked by implementation issues including low individual take-up and delays in program implementation.

Aware of the shortcomings of the first early childcare expansion, the Government of Cambodia, with the support of the World Bank, launched another preschool education expansion program for the period 2014—2018. The plan was financially supported by a second GPE grant (GPE II) of \$38 million, of which about \$20 million were allocated to ECD programs. Our study focuses on the sub-component that includes the construction of formal community preschools.

2.2 Supply-side intervention: Formal community preschools

Before GPE II, two distinct types of public preschools existed in Cambodia: state preschools (SPS) and “community preschools.” Since these “community” preschools were unstructured and lacked uniform quality standards, we refer to them here as informal (community) preschools (IPS).^{6,7} GPE II introduced a new type of community preschool with a structured age-appropriate educational program and uniform quality standard, which we refer to as (formal) community preschools (CPS).

SPS were financed by the Ministry of Education, Youth and Sport (MoEYS). SPS teachers benefited from two years of formal training in a MoEYS teacher training center in Phnom Penh. They received a monthly salary of roughly \$250 to teach for three hours a day, five days a week. Almost all SPS were attached to a public primary school and had access to classrooms equipped with teaching and play materials, along with better overall infrastructure (including sanitation facilities).⁸

In contrast, IPS were typically not attached to a primary school. The local community established an IPS and covered operational costs. This included the IPS teacher stipend, which was at the discretion of the local commune council. Constrained by council budgets, it varied from \$30 to \$50 per month at the time of

⁵Source: Data from UNESCO Institute for Statistics.

⁶According to government data (MoEYS, 2017), of the 7,241 preschool facilities in Cambodia in 2016, 55% were SPS, 39% were IPS, and 6% were private preschools. However, these preschools are not evenly distributed across the country, and 38% of the 1,646 communes in Cambodia had no preschool facility.

⁷See Bouguen et al. (2013) for an impact evaluation of each type of preschool developed in the context of the GPE I.

⁸The program designers and research team did not anticipate the presence of SPS in the experimental villages. Villages with a preexisting SPS were not supposed to be eligible for the program.

our baseline; most IPS teachers relied on other income sources for their livelihood. IPS teachers were trained for about 35 days by provincial education departments before they began work. Teachers provided a two-hour preschool class five days a week. The quality of IPS could differ substantially across villages as communes established IPS with their own funds. IPS classes were often held in a teacher's home, a community hall, or a pagoda.

To increase preschool access and to improve the unsatisfactory quality of IPS, the Cambodian government used the GPE II grant to establish 500 new CPS. Most of these replaced previously existing informal arrangements (i.e. the IPS were shut down when the CPS was built); some were established in villages that previously had no preschool or were too large to be served by one preschool alone. Only one CPS was established in each eligible village and anecdotal reports suggest that when a CPS replaced an IPS, the IPS teacher became the CPS teacher.⁹ In contrast to poorly resourced and low-quality IPS, CPS benefited from a standardized building, fully equipped with tables, chairs, and a blackboard (See the pictures in Figure A1), directly financed by the GPE II. Each CPS had a capacity of 25 children.

While the preschool curriculum was similar in IPS and CPS, the CPS teachers received more training: Their training also lasted 35 days but it included structured lessons in pedagogical strategies, curriculum content, testing, and how to operate a CPS. Teachers were also trained in the basics of child development, child rights and parental education. All teachers participated in a written examination before and after the training. Further, teachers were provided with a package of teaching materials tailored to the CPS curriculum. Like IPS teachers, CPS teachers were usually a community member who, after completion of the training, provided a two-hour class each day, five days a week, to children aged three to five years. Their salary was similar to IPS teachers.

The Cambodian government considered CPS a promising, affordable alternative that could prove similarly effective to SPS. CPS require fewer resources for building construction and teacher salaries than SPS, since teachers are relatively less educated and are recruited locally. A similarly quick and large scale-up of SPS would have been significantly more costly (and likely unfeasible). It also would have proven difficult due to the more intensive teacher training program and the lack of sufficient personnel to train many teachers at once.

2.3 Demand-side intervention and parenting program : Door-to-door awareness campaign and home-based program

The door-to-door (D2D) program implemented as a part of this intervention was a demand-side intervention aimed at stimulating demand for early childcare programs by speaking directly to individual caregivers. The goal was to sensitize them to the

⁹We do not have the data to confirm this.

value of preschool education and guide them through the enrollment of their children at CPS. An additional component was to provide information about returns to education. Such information has been shown to effectively increase attendance and change the social composition of students in other lower-income contexts (Nguyen, 2008; Jensen, 2010). The local village head and the field staff responsible for the study’s data collection performed the D2D activities.

At baseline, caregivers were informed about the new preschools and key details, such as that a preschool has been constructed and that there was no school fee. In addition, caregivers received a printed leaflet that had more information about the newly established CPS (see Figure A2). It noted that the teacher had been trained and it suggested how preschool in general could help children improve primary school readiness and, potentially, their overall educational attainment. The leaflet also provided information about average incomes in Cambodia by educational background, visualized in a graph using data from the Cambodian Socioeconomic Survey 2009. Caregivers received another informational leaflet about one year later, at midline.¹⁰ Importantly, since survey field staff were responsible for the leaflet distribution (after the baseline and midline survey), we can ascertain that all eligible respondents received at least the information contained in the leaflet.

The home-based program (HBP) formerly operated in Cambodia as an independent early childhood education service to support parents of children aged zero to five years. However, it was redesigned as supplementary to CPS, aiming to enhance the effect of CPS enrollment. The program was implemented by local “core mothers” who received initial and ongoing training from MoEYS. The 35 day long training covered a wide range of subjects such as child rights, pre- and postnatal care of mothers, hygiene, nutrition, disease prevention, developmentally appropriate activities for children, school readiness, disabilities, health services, and child protection. Similarly to CPS teachers, core mothers participated in a written examination before and after the training. They were responsible for promoting enrollment of children aged three to five years into CPS and for leading monthly informational meetings with parents of children aged zero to six. Core mothers were volunteers who only received stipends while in training. The HBP was supposed to take place regularly; it was designed as a more intensive demand-side intervention than the D2D as well as a light supply-side intervention, complementary to CPS.

¹⁰The content of the leaflets was developed in cooperation with MoEYS. After baseline, we received feedback from village heads that the leaflet contained too much information (see Figure A2). Therefore, the midline leaflet was simplified to focus only on the advent of a new CPS (see Figure A3).

2.4 Costs of CPS and other preschools

Our cost estimates are based on the ingredient method¹¹ which aims to cost every resource required to make an intervention happen.¹² This includes construction costs as well as support costs, such as management, administrative and overhead costs. Total annual costs for running 500 CPS is estimated to be \$3,136,743 for an average cost per school of \$6,273 or \$256 per child, which compares to roughly 22% of the GDP per capita (\$1,160 in 2015). The estimated annual SPS costs are much higher at \$12,602 per school and \$426 per child.¹³ Due to the lack of uniform quality standards no cost estimates for IPS were calculated. However, we anticipate the cost of IPS to be vastly lower. We obtain a back-of-the-envelope cost estimate of the IPS by using the CPS cost analysis from which we subtract all costs related to the construction (3%), the material (25%), the equipment (4%).¹⁴ Based on this approach, the cost of IPS is estimated to be 68% of that for CPS i.e. \$174.

3 Evaluation Design and Data

3.1 Randomization and sampling

This evaluation of the CPS program is based on a randomized controlled trial. All sample villages are situated in the south and northeast parts of Cambodia. Eligibility criteria for villages to participate in the study were: expressed demand for a CPS, a high poverty rate, and a large number of children aged zero to five years.

The study sample is composed of 305 villages. Before baseline, villages were randomly assigned to the control group or one of the three treatment groups: CPS (T1); CPS and D2D (T2); or CPS, D2D, and HBP (T3). Randomization was stratified to obtain a sample for which treatment is balanced within each of the 13 provinces of our sample.¹⁵ The design is summarized in Table 1.

¹¹Therefore, our cost measures are not directly comparable to studies that base their cost-effectiveness analyses on programmatic costs only (e.g. Brinkman et al. (2017))

¹²A full discussion of cost estimates can be found in the SIEF impact evaluation report (Berkes et al., 2019).

¹³The cost estimates reported here are based on a model using mid-range estimates for unobserved costs. Under all models, costs per child at SPS were 50%—126% higher than at CPS.

¹⁴This is probably an overestimation of the IPS cost since IPS cater to more students and the training costs were higher for CPS than they were for IPS.

¹⁵The randomization was performed with a list of 310 eligible villages provided by the government. Of these, 60 were assigned to the control group, 123 to T1, 63 to T2, and 64 to T3. Unfortunately, the randomization list contained erroneous village names and five of them were duplicated or could not be identified after the randomization even after substantial effort by MoEYS and the data collection firm. Therefore, the total number of villages decreased to 305. We treated this dropout as random and did not replace the villages. The randomization list also contained villages for which a CPS teacher was no longer available or for which no land could be secured for CPS construction. Therefore, only 91% of treatment group villages received a CPS. Since these factors are potentially endogenous, we do not treat these as random. To maintain ex-ante expected balance between control and treatment villages, these villages were therefore not removed from our sample.

Table 2 gives an overview of data collection activities and timing of the preschool construction. Our analysis is based on three main waves of data collection: a baseline data collection in May—July 2016, a midline survey in April—June 2017, and an endline survey in May—July 2018.¹⁶ With 82% of CPS completed for the first school entry in October 2016¹⁷ and 91% of CPS constructed at endline, Table 2 confirms the construction rolled out as planned. Nevertheless, and despite our effort to ensure that the baseline survey preceded CPS construction, a completed CPS building was already available at baseline in 17% of the treatment group villages. It is challenging to conduct an experiment like this with school construction. On the one hand, fielding the baseline too early (well before any construction) would have increased the risk (in case of construction delay) that our baseline sampled children would have been too old to attend the newly built preschools.¹⁸ On the other hand, fielding the baseline too late would have resulted in baseline measures that were arguably already affected by the program. We discuss the implications of the slight overlap between baseline survey and construction below.

During baseline data-collection in 2016, we sampled up to 26 eligible households per village, using an adapted version of the EPI walk,¹⁹ which guarantees representativity of households with children of preschool age.²⁰ Eligible households include at least one child between 24 and 59 months old at baseline. Children were therefore between three and five years old at midline and between four and six years old at endline. We identified 7053 eligible households at baseline. Using the same sampling method, we added extra households at midline in villages where the number of eligible households at baseline was below 10 (53 were added in total).

3.2 Data

3.2.1 Survey and instruments

For each household, a household survey, caregiver survey, and one assessment per eligible child were conducted.²¹ At the village level, interviews were conducted with village heads and preschool teachers. In addition, at endline, we complemented the

¹⁶A brief monitoring survey was conducted in late 2016 to confirm that CPS construction was proceeding as scheduled.

¹⁷17 CPS were in construction at that date. 10 of them would be completed by the end of the calendar year 2016. It further means that the disturbance caused by the construction is probably minimal

¹⁸As described in Bouguen et al. (2013), construction delays occurred in a previously evaluated program in Cambodia. This considerably reduced take-up, exposure time, and statistical precision.

¹⁹EPI refers to the Expanded Programme on Immunization of the World Health Organization; see e.g., Henderson and Sundaresan (1982).

²⁰The sample is not nationally representative. Villages were selected based on criteria such as expressed interest in the program, poverty rate, lack of a functioning preschool, and teacher availability. Sample households are exclusively from rural areas in southern and eastern Cambodia. This is because western provinces received CPS under a previous preschool construction program.

²¹The caregiver is defined as the direct relative who takes care of the child most of the time. In most cases, the caregiver is a biological parent (60.4% at baseline).

data collection with a classroom observation survey conducted in all the preschools within the sample villages.

The household survey included information about family structure, household wealth, and other socioeconomic background characteristics. The caregiver survey included questions regarding the child’s preschool enrollment and socio-emotional development, as well as 25 questions about parenting practices. The latter measure three key dimensions of parenting: “cognitive parenting”, “emotional parenting”, “and negative parenting”. More cognitive parenting means parents are more likely to engage in activities that contribute to the development of their child’s cognitive competencies (e.g. by playing games, reading books, or playing with toys). More socioemotional parenting means parents interact more in ways that provide emotional support and responsiveness (e.g. comforting, encouraging or complimenting the child). More negative parenting means parents are more likely to use harsh or punitive approaches to discipline their child. Last, the survey included information about home visits, D2D activities, and participation in the HBP, as well as the caregiver’s perceived returns to education, and a short test of the respondent’s nonverbal reasoning ability (based on the Raven’s Progressive Matrices Test).

The approximately 45-minute child assessment included a comprehensive battery of cognitive tests (executive function, language, early numeracy, fine motor development and, at baseline and midline only, gross motor development) as well as anthropometric measures (height and weight). Most of the child tests were based on the Measuring Early Learning Quality and Outcomes (MELQO) toolkit. (See UNESCO (2017) for a description of the measures-of-development process and Raikes et al. (2019) for evidence of validity).²²

Additional child tests were added to the MELQO items to increase the sensitivity and breadth of the child assessment. The additional tests included the following: the Dimensional Change Card Sort (Zelazo, 2006)²³; a receptive vocabulary test based on picture recognition; a test for knowledge of reading concepts (based on a monitoring tool used by the Cambodian Ministry of Education); and a sustained attention test. Children’s socio-emotional development was measured using the caregiver-reported Strengths and Difficulties Questionnaire (SDQ). Using these tests, we create two indexes: a cognitive development index which aggregates executive function, language, early numeracy, and fine motor scores and a socio-emotional index which aggregates the individual dimensions of the SDQ.²⁴

²²An in-depth discussion of midline child tests, scoring methods, cultural adaptations, pretesting procedures, and questions about parenting practices can be found in Berkes et al. (2019).

²³No cultural adaptation of words or pictures was conducted since the test was working well in field pilots and changes were not deemed necessary by the local staff.

²⁴Individual tests were scored and standardized to ensure that their variance equally contribute to the composite score. For almost all tests, we assign one point per correct answer and zero when the child was unable to complete the practice trial of a test. We calculated each test score by summing all correct answers and standardizing them using the control group sample mean and the control group sample standard deviation. All standardized test scores of one domain (e.g.,

The village and teacher surveys included questions about the ECD services available in the sample village. This allowed us to monitor implementation of program interventions and precisely know when each preschool service was available. Parts of the teacher survey and the classroom observation tool administered at endline are based on the Measuring Early Learning Environments (MELE) module of MELQO. MELE includes key domains for quality in early learning environments and the sample items used to measure them. During pretests preceding the endline data collection, constructs to be measured were selected and specific items were adapted to take into account culture-specific views on what defines a high-quality learning environment. We divided the final module into five domains: teacher characteristics, equipment, classroom setting, curriculum content and pedagogy, and teacher-child interactions.

3.2.2 Balance and Attrition

The main baseline characteristics of villages, households, children, and their caregivers are summarized in Table 3. To test for statistically significant differences, we regress each variable on binary indicators for the treatment groups and a set of province dummies to account for stratified randomization. Overall, the sample is balanced in child, household, and caregiver characteristics.

Table 3 also highlights some significant baseline differences in preschool enrollment and take-up of demand-side interventions. These differences are due to the aforementioned early roll-out of some preschool interventions. Children in program villages are about 8 pp more likely to have enrolled in preschool at baseline (driven by community preschools which, by definition, were newly constructed). Similarly, the HBP intervention started in some treatment villages before baseline. These baseline imbalances at the individual level are confirmed by imbalance in the availability of preschools and HBP at the village level, shown in the baseline panel of Table 5. At baseline, some villages in the treatment group had access to CPS and many treatment groups' villages had access to HBP. There are reasons to believe that these early interventions are unlikely to affect our estimation. baseline results do not pick up any significant cognitive differences between the treatment groups and the control, indicating that early enrollment in preschool intervention had not given a significant benefit to the treatment group. Relatedly, expressed in number of days at school, early enrollment in preschool programs is only marginally higher in the treatment groups. At baseline, treatment children had approximately only 10 extra days of preschool compared with the control. It further means that the roll-out of our baseline survey was on average 10 days too late compared to the roll-out of the the CPS construction. Given the size of the program, we consider

executive function) were then summed into a domain score and standardized again. The individual tests and scoring methods are summarized in Appendix 6.

this small 10-day discrepancy to be acceptable and unlikely to drive results. The case is less clear for HBP, which had been rolled out before baseline.²⁵ Since HBP is mainly a demand-side intervention with limited expected short-term effects on cognitive performances, early participation of parents in the HBP is unlikely to have undermined to our experimental design.

A potentially greater cause for concern is the fact that SPS are reported, at baseline and later on, to be less available in the treatment groups compared to the control group (Table 5). While the point estimates are insignificant for T1 and T3, and again insignificant when the treatment groups are considered jointly (- 7 pp), the T2 villages are significantly less likely to have access to an SPS (-15 pp). Since SPS are expected to be of much better quality than CPS (as we discuss in section 3.2.3 below), this small imbalance, which we attribute to chance, is likely to provide a baseline (relative) cognitive advantage to the control group, therefore driving down our estimates of CPS impacts on cognitive development. This downward bias is likely to be stronger in T2. We return to this then discussing results.

Finally, with regards to attrition, we find no correlation between the random assignment and the probability of not responding both at midline and endline (Column (1) in Table B1). Attrition is not significantly related to a set of variables strongly associated with child development (Column 2). Furthermore, there is also no evidence of differential attrition with respect to baseline controls except for T2 at the endline follow-up, where children with a high height-for-age z-score were more likely to attrit from the sample. The addition of households to the sample at midline causes differential attrition to become significant for T1 at midline since fewer households were added to this group (Column 3). The overall level of child attrition is around 10% at both midline and endline, and mostly due to seasonal migration.

3.2.3 Quality of CPS and other preschools

We construct measures of preschool quality based on data from two sources, classroom observations and parent reports.

Preschool Quality Measures

We use the teacher survey and classroom observation tools to assess how new CPS compare to SPS and IPS in terms of structural and process quality. We compare the different preschools in three dimensions of structural quality (teacher characteristics, equipment, classroom setting) and two dimensions of process quality (curriculum and pedagogy, teacher-child interactions).

Individuals quality measures are summarized using aggregate measures for each of the five dimensions of structural and process quality, calculated as the first principal component of individual variables (Table 4, panel “Preschool comparison”).

²⁵This is probably due to the fact that setting up an HBP takes much less time than building a CPS.

CPS and IPS teachers are similar according to the teacher characteristics score (Column 4), while SPS teachers perform significantly better in this dimension (+1.7 SD column 5). However, CPS outperforms IPS on the classroom setting and equipment scores. In these two dimensions, CPS are equivalent and maybe even superior (for equipment) to SPS, confirming that structural quality has been significantly improved by the construction program. The results are strikingly different for the process quality scores. CPS and IPS are not significantly different in terms of curriculum and pedagogy or teacher-child interaction, whereas on both aggregate measures, SPS outperform IPS and CPS.²⁶

Overall, these results confirm that CPS are of better quality than IPS but maybe not on the dimensions that are the most important for children’s performance. The upgrading from IPS to CPS mostly affected the learning environment by providing better equipment, while teacher quality was not fundamentally affected. In contrast, SPS teachers are significantly more educated and perform better in terms of instructional quality.

Parental assessments

We use questions from the caregiver survey to compare differences in cancelled classes, perceived teacher quality, financial contributions, and travel time between the types of preschools (Table B4). CPS classes get cancelled significantly more often than SPS classes. Parents of children enrolled at CPS assess the kindness of CPS teachers as higher than that of their SPS counterparts (though these differences tend to be relatively small). This could be explained by the greater likelihood of CPS teachers being from the same community as the parents. At endline, parents of children enrolled at SPS assess the reliability of SPS teachers as higher than that of CPS teachers. While financial contributions for teacher salary and construction tend to be almost zero for all types of schools, contributions for school materials are substantial: On average, parents of children enrolled at CPS have paid \$35 in the current school year at midline and \$47 at endline (with similar amounts for IPS: \$32 at midline, \$53 at endline); contributions at SPS were \$49 at midline and \$68

²⁶Unpacking these comparisons into the individual items that make up the aggregate scores shows that SPS teachers are more experienced, perform better in the Raven’s Progressive Matrices Test and have more training than IPS and CPS teachers (Tables B2 and B3). They are also better paid and receive their salary more regularly. In terms of equipment, CPS significantly dominate IPS and even in some dimension SPS, except for access to electricity, water source and toilet facilities. Compared with IPS, CPS classrooms have less breaks, are more likely to follow a curriculum, have more attendance, more enrollment, have more storybook activities (Table B3). Yet, the quality of the classroom setting, pedagogy and teacher-child interactions remain far from the quality observed in SPS. Crucially, the teacher-interaction individual variables are very similar in CPS and IPS, while SPS engage more with the children and provide significantly more encouragements.

at endline.²⁷²⁸

4 Empirical Strategy

4.1 Reduced-Form estimation

By virtue of randomization, we can straightforwardly estimate ITT impacts of the intervention:

$$Y_{iv} = \alpha_0 + \alpha_1 Z_v + \mathbf{W}'_{ip} \alpha_2 + \mu_v + \epsilon_{iv}, \quad (1)$$

where Y_{iv} denotes the outcome of child i in village v , Z_v the treatment assignment in the pooled specification—replaced by a vector of three binary indicators for treatment groups T1, T2, T3, and α_1 by a vector of corresponding treatment effects in the disaggregated specification—and \mathbf{W}_{ip} a set of control variables. μ_v and ϵ_{iv} are the unobserved village- and individual-specific error term components, both assumed uncorrelated with \mathbf{W} and Z . We use standard errors clustered at the village level to account for the randomization implemented at that level. \mathbf{W}_{ip} includes the baseline child test score, child age quarter dummies, and gender in the basic specification. The specification also includes μ_p , which is a set of province fixed effects to account for stratified randomization within provinces. (Bruhn and McKenzie, 2009)

The ITT estimate is empirically valid and policy relevant, as it reflects the program’s impacts in a typical Cambodian preschool environment. Yet, in our context, the ITT effects should be interpreted with the understanding that they stem from children who would not have benefited from any preschool absent the program (*h-compliers*), but also from children switching from an existing preschool arrangement (either IPS, SPS or primary school²⁹ to the CPS (*a-compliers*)). The latter substitution patterns may mask potentially larger impacts on *h-compliers* which are policy-relevant in their own right. In addition, estimating the program’s impacts on children who would have attended an alternative preschool (*a-compliers*) is relevant to measure how successful the CPS program was at improving preschool quality in this context.

²⁷The exact question asked is “How much money have you spent on school material (paper, board, chalk, cloth, water, food) for your child since the beginning of the school year?” Hence, it does not distinguish between money for food that is used for “pedagogical” reasons and money that is used to buy snacks for consumption. While we do not have data on this, we cannot rule out that teachers sell food as an indirect way of increasing their salaries.

²⁸At most CPS and IPS, the school year starts in late October or early November. About half of CPS and IPS end their school year at the end of July, while the other half end the school year at the end of August. Most SPS have their school year from late October or early November until the end of August. Since the endline data collection was conducted about one month later than the midline data collection, some caution is warranted with comparisons between the waves.

²⁹Primary school enrollment is almost zero at midline but is about 25% at endline.

4.2 LATE and subLATEs

The LATE, that we will refer to as $LATE_{cps}$, can be estimated as usual by instrumenting CPS enrollment ($\mathbb{1}_{\{D_i=c\}}$) with Z_v , with D_i taking value c, a or h depending on whether the child is enrolled in a CPS (c), in an alternative school program (or ASP which regroups IPS, SPS or primary school) (a) or is staying at home (h). $\mathbb{1}_{\{D_i=c\}}$ is a dummy variable taking value one when the child is enrolled in CPS and zero otherwise. Just like the ITT parameter, the $LATE_{cps}$ is identified but should be interpreted as a weighted average between the impact of children who would have stayed at home—which we refer to as the subLATE on *h-compliers* or $LATE_{hc}$ —and the impact on children who would have attended an alternative school—the subLATE on *a-compliers* or $LATE_{ac}$. As discussed by Kline and Walters (2016), $LATE_{cps}$ can be decomposed into of the two sub-LATEs :

$$LATE_{cps} = S_{ac}LATE_{ac} + (1 - S_{ac})LATE_{hc} \quad (2)$$

where S_{ac} , the share of *a-compliers* (within the group of compliers), is identified and given by Kline and Walters (2016):

$$S_{ac} = \frac{P(D = a|Z = 0) - P(D = a|Z = 1)}{P(D = c|Z = 1) - P(D = c|Z = 0)} \quad (3)$$

Figure B1 provides a visual representation of the parameters in Equation (2) for groups C and T1: the share of *a-compliers* is visually represented by the *a-compliers* region divided by the region occupied by any compliers, and $LATE_{cps}$ is a weighted average of both sub-LATEs, $LATE_{ac}$ and $LATE_{hc}$.

4.3 Identifying Bounds for $LATE_{hc}$

Equation (2) makes explicit the challenges faced in estimating the impact of a policy in a context of close substitutes. Under traditional assumptions (such as those discussed in Imbens and Angrist (1994)), $LATE_{cps}$ is identified, but its sub-LATE components ($LATE_{ac}$ and $LATE_{hc}$) are not.³⁰ However, under plausible assumptions, we can derive bounds for the $LATE_{hc}$, with the key assumption being:

$$0 \leq LATE_{ac} \leq LATE_{hc} \quad (4)$$

i.e., the CPS offer a better learning environment than the average alternative school program (ASP) (left hand side of the inequality) and that *h-compliers* benefit more

³⁰The assumptions needed to secure the LATE are slightly modified in presence of alternative programs, specifically the exclusion restriction and the monotonicity assumptions needs to account for the presence of the alternative program.

from enrolling to CPS than *a-compliers* do (right hand side of the inequality).³¹

The left hand side of inequality (4), $0 \leq LATE_{ac}$, simply implies that switching from an ASP to a CPS is not, on average, detrimental to *a-compliers*. Given the resources devoted to CPS in comparison with IPS, we believe that the left hand side of (4) inequality is very likely to hold.³² The right side of inequality (4), $LATE_{ac} \leq LATE_{hc}$, implies that *h-compliers* benefit from a greater improvement in their learning environment than *a-compliers*. Since *a-compliers* already benefit from some preschool intervention regardless of their treatment status, this assumption is intuitively very likely.

Under (4), we obtain $LATE_{hc}$'s lower bound by inserting the right side of inequality (4) into (2):

$$LATE_{ac} = LATE_{hc} \iff LATE_{hc}^{LB} = LATE_{cps} \quad (5)$$

i.e. the low bound assumes that *h-compliers* and *a-compliers* benefit equally, on average, from the CPS intervention, hence, $LATE_{cps}$ is the low bound. Using the left side of inequality (4) we obtain the upper bound:

$$\begin{aligned} LATE_{ac} = 0 \iff LATE_{hc}^{UB} &= \frac{LATE_{cps}}{(1 - S_{ac})} \\ &= \frac{\beta_1^{ITT}}{\beta_1^{FS_c} * (1 - S_{ac})} = \frac{\beta_1^{ITT}}{\beta_1^{FS_c} * (1 + \frac{\beta_1^{FS_a}}{\beta_1^{FS_c}})} \quad (6) \\ &= \frac{\beta_1^{ITT}}{\beta_1^{FS_c} + \beta_1^{FS_a}} = \frac{\beta_1^{ITT}}{\beta_1^{FS_{ps}}} \equiv LATE_{ps} \end{aligned}$$

with $LATE_{cps} = \frac{\beta_1^{ITT}}{\beta_1^{FS_c}}$, $\beta_1^{FS_c}$ the CPS first stage parameter, $\beta_1^{FS_a}$ the equivalent first stage parameter for the ASP, $\beta_1^{FS_{ps}} = \beta_1^{FS_a} + \beta_1^{FS_c}$ the first stage parameter that captures the differential *any* preschool take-up, and $LATE_{ps}$ the effect of *any* preschool enrollment instrumented by Z. Essentially, our arguments imply:

$$LATE_{cps} \leq LATE_{hc} \leq LATE_{ps}$$

$LATE_{hc}$ is bounded by $LATE_{cps}$ and $LATE_{ps}$.

Following a similar approach as (Lee, 2009), we can narrow the bounds using a set of additional variables orthogonal to Z. We assume:

$$0 \leq LATE_{ac}(\mathbf{B}) \leq LATE_{hc}(\mathbf{B}) \quad (7)$$

³¹ASP encompasses students enrolled in IPS, SPS or primary school

³²As we will see later (see Section 5.2, the substitution is essentially occurring between IPS and CPS as the construction of a CPS forces the IPS to shut down. Yet, the SPS enrollment is also slightly affected. We argue that this small difference is indeed entirely due to baseline imbalance, not a substitution between school offer.

which is the equivalent to equation (4) for each value of \mathbf{B} , the variables orthogonal to \mathbf{Z} . We can implement the bounding strategy – calculate $LATE_{cps}$ and $LATE_{ps}$ for each value of \mathbf{B} – in the sample cells formed by \mathbf{B} if \mathbf{B} is categorical. We then average across the values of \mathbf{B} to recover the unconditional narrow lower and upper bounds, using the probability of belonging to each cell as weights. The lower bound can also be estimated more flexibly using the following IV regression model:³³

$$Y_{iv} = \beta_0^{lb} + \beta_1^{lb} \mathbb{1}_{\{D_i=c\}} + \mathbf{B}_i \beta_2^{lb} + \mathbf{W}_i \beta_3^{lb} + u_{iv} \quad (8)$$

where $\mathbb{1}_{\{D_i=c\}}$ – the dummy for CPS enrollment – is instrumented by \mathbf{B} , \mathbf{Z} , $\mathbf{Z} * \mathbf{B}$ and controlling for \mathbf{W} . β_1^{lb} is the parameter of interest that gives the narrow lower bound.

Similarly, the narrow upper bound can be estimated using:

$$Y_{iv} = \beta_0^{ub} + \beta_1^{ub} \mathbb{1}_{\{D_i=a \cup D_i=c\}} + \mathbf{B}_i \beta_2^{ub} + \mathbf{W}_i \beta_3^{ub} + u_{iv} \quad (9)$$

where $\mathbb{1}_{\{D_i=a \cup D_i=c\}}$ – the dummy for any preschool enrollment (CPS or ASP) – is instrumented by \mathbf{B} , \mathbf{Z} , $\mathbf{B} * \mathbf{Z}$ and controlling for \mathbf{W} . While any baseline variable can theoretically be included in \mathbf{B} , the choice of the \mathbf{B} variables depends on two potentially conflicting criteria. First, the size of the narrow bounds will depend on the ability of the \mathbf{B} variables to predict the enrollment behavior. Second, the \mathbf{B} variables should be sufficiently parsimonious to maintain a reasonable sample size in each cell (and therefore statistical precision). Also, assumption (7) needs to hold in each cell formed by \mathbf{B} , an additional reason to limit the number of cells formed by \mathbf{B} . To balance both criteria, we estimate the narrow bounds using the median of the ASP enrollment’s predicted value as defined below in Section 4.5, which allows both a reasonable sample size in both cells and a good prediction of the enrollment behavior.³⁴

4.4 Estimating the sub-LATE using Conditional LATE

In addition to not being able to pinpoint the exact magnitude of the $LATE_{hc}$, the bounding approach has one additional limitation in that it does not allow for an estimate of $LATE_{ac}$. While $LATE_{hc}$ is arguably a more central parameter to the preschool literature (that is, this new preschool versus no preschool), it is also important to assess to what extent the newly constructed school is an improvement over the current provision of education in this particular context. If $LATE_{ac}$ is zero

³³To see this, let \mathbf{B} be a dummy variable taking two values. To calculate the narrow lower bound, we jointly estimate an IV regression for $\mathbf{B}=0$ and $\mathbf{B}=1$ and we take the weighted average using the probability of belonging to group $\mathbf{B}=1$ and $\mathbf{B}=0$ respectively. Doing so corresponds to an IV regression where $D_i = c$ is instrumented by \mathbf{Z} , \mathbf{x} and $\mathbf{X} * \mathbf{Z}$.

³⁴we create a dummy taking value 1 when the ASP enrollment’s predicted value is above the median

or close to zero, this would suggest that the new CPS did not fundamentally improve the available provision of education.

To identify the subLATE parameters, we would ideally need an extra source of variation that affects the share of *a-compliers* (within the compliers) without directly affecting the CPS treatment effects. These additional instruments (which we refer to as \mathbf{X}) should affect the treatment effects only through the share of *a-compliers* and not because they capture other forms of heterogeneity (e.g. heterogeneity driven by gender). This assumption is called constant treatment effect (Kline and Walters, 2016) or homogeneous LATE (Hull, 2018). We refer to this approach as conditional sub-LATE. Kline and Walters (2016) (see also Hull (2018) and Feller et al. (2016)) show that sub-LATEs can be identified by interacting the random assignment with observed covariates. The structural equation takes the following form:

$$Y_{iv} = \gamma_0 + \gamma_1 \mathbb{1}_{D_i=c} + \gamma_2 \mathbb{1}_{D_i=a} + \mathbf{X}_i \boldsymbol{\gamma}_3 + \mathbf{W}_i \boldsymbol{\gamma}_3 + u_{iv} \quad (10)$$

where Y_i is a follow-up outcome, $\mathbb{1}_{D_i=a}$ a variable taking one when the child i is enrolled in an ASP and zero otherwise, \mathbf{X} a set of additional variables orthogonal to the treatment, and \mathbf{W} the preferred set of control variables used for the reduced form estimation. γ_1 captures the $LATE_{hc}$, and γ_2 captures the effect of going to ASP. To derive the $LATE_{ac}$, we subtract γ_2 from γ_1 . $\mathbb{1}_{D_i=c}$ and $\mathbb{1}_{D_i=a}$ are both endogenous and instrumented using following first stage equations:

$$\begin{aligned} \mathbb{1}_{D_i=c} &= \pi_0^c + \pi_1^c Z_v + \pi_2^c Z_v * \mathbf{X}_i + \mathbf{X}_i \boldsymbol{\pi}_3^c + \mathbf{W}_i \boldsymbol{\pi}_4^c + \mu_v + \epsilon_{iv} \\ \mathbb{1}_{D_i=a} &= \pi_0^a + \pi_1^a Z_v + \pi_2^a Z_v * \mathbf{X}_i + \mathbf{X}_i \boldsymbol{\pi}_3^a + \mathbf{W}_i \boldsymbol{\pi}_4^a + \phi_v + \nu_{iv} \end{aligned} \quad (11)$$

The identification of γ_1 and γ_2 relies on the independence of Z and $Z * \mathbf{X}$ and on the assumption that the h and a compliers have a constant return to preschool on \mathbf{X} . Since the \mathbf{X} variables are selected from the baseline survey, their independence is guaranteed by randomization. However, these extra instruments could capture sources of variation that are not related to the variation in counterfactual enrollment, therefore causing a violation of the constant treatment assumption. To be valid, the conditional LATE approach therefore requires that the extra instruments capture the heterogeneity caused by the variation in counterfactual enrollment (i.e. the fact that some would have stayed at home and some would have enrolled in an ASP) but not other forms of more standard heterogeneity. Arguably, variables at the province or village level that capture infrastructure quality are less likely to be correlated with the standard heterogeneity associated with observed characteristics. Note that when there is more than one instrument in \mathbf{X} , the validity of the constant treatment assumption can be tested using an over-identification test.

4.5 Choosing the right instruments

The validity of the conditional LATE approach—and to a lesser extent of the narrow bounds approach—crucially depends on the choice of instruments. We adopt two strategies for this choice. First, we use an *ad hoc* strategy in which we select a number of variables at the village and at the household level likely to affect the share of *a-compliers* (within the group of compliers). For instance, we include the province fixed effects under the expectation that some provinces may have provided, before the beginning of the intervention, better preschool infrastructure than others without necessarily affecting the treatment effects. Using the same logic we include population size and land area of the villages³⁵ and whether the village has a primary and/or a secondary school. In a second specification within this same strategy, we add caregiver level variables (education and poverty level as well as caregiver cognitive test scores).

In a second strategy, we improve the transparency of the variable selection by implementing a machine learning algorithm to determine which variables should be included in equation (10) as X instruments to predict ASP attendance. We leverage the large baseline dataset composed of variables at the province, village, household, caregiver and child levels. Our objective is to predict the probability that a child will be enrolled in an ASP using baseline variables following the procedure implemented in Belloni et al. (2013) and Chernozhukov et al. (2018). We start the procedure by selecting all covariates available at baseline (1354 in our case). We exclude free-text variables, variables with very low response rate,³⁶ and variables without variation, which results in 224 eligible variables. We then create dummy indicators for categorical variables, compute the square of each covariate, and then remove perfectly collinear variables.³⁷ We use the resulting 493 baseline variables to predict ASP enrollment status at follow-up in the control group using LASSO.³⁸ To select the penalty term, we compare three different approaches: cross-validation, adaptive (Zou (2006)), and plug-in (Belloni et al. (2014)). Comparing the postestimation deviance ratio, we conclude that the plug-in LASSO fares slightly better and is much more parsimonious: in addition to the 12 province fixed effects, the plugin LASSO identifies five baseline variables that best predict midline ASP enrollment in the control group (whether the child attended preschool at baseline, whether he/she attended more than 30 days, baseline cognitive scores, age squared and one variable taking value one when the baseline infrastructure information is missing).³⁹

³⁵Under the expectation that population density may have an effect on alternative preschool availability

³⁶We exclude variables with response rate lower than 70% in the control group with non missing ASP attendance information.

³⁷To avoid dropping too many valuable observations, we impute baseline missing values using the median variable value. When we impute values, we create a dummy variable that takes value 1 for the imputed observation and 0 otherwise.

³⁸We partial out the LASSO with stratification variable i.e. province fixed effect.

³⁹The LASSO that uses the data-dependent penalty developed by Belloni et al. (2012) (some-

We use the resulting variables to predict the propensity to attend ASP, \hat{D}_a , using a logit model. We then use \hat{D}_a as instrument in equation (10) and use the median \hat{D}_a as a dummy variable to narrow the bounds in equations (8) and (9).

The LASSO algorithm has two advantages compared to our first *ad hoc* approach: First, the LASSO algorithm is more transparent. Second, the LASSO is designed to select the instruments that best predict ASP enrollment, therefore increasing the precision of the estimation. But the approach has one important drawback: it may choose instruments that are related to other forms of heterogeneity. In the *ad hoc* method, we alleviate this concern by excluding variables at the child level.

5 Results

5.1 Impacts on childcare provision

Before discussing the impact of the intervention on outcomes, we review the CPS construction status and overall availability of preschools, and quality. As discussed above, there are imbalances in SPS availability at baseline (-14.5 pp) that remain observable at midline (-13.0 pp) and endline (-13.2 pp) (Table 5). We attribute these imbalances to chance and note that, if anything, such imbalance should drive our estimates of impacts on child development downward. In addition, since the imbalance is larger in T2, the downward bias should also be larger in T2, a hypothesis that we return to in Section 5.2.

Importantly for interpreting our results, 81% of control group villages have some sort of preschool at midline; by endline, this is almost 85%. In the treatment groups, preschool availability is close to 100% already at midline and most villages (80%) only have one preschool. As mentioned, most formal CPS replaced preexisting alternative preschools and most of this replacement happened between CPS and IPS. For instance, 96% of the treatment villages benefited from a CPS by midline and, at the same time, 57% of them saw their IPS closing while the number of SPS is barely affected (-8%).⁴⁰ Since we expect CPS to be of better quality than IPS, this substitution pattern may suggest that preschool quality has improved in the treatment groups.⁴¹ We use the quality measures presented in Table 4 (panel “Experimental comparison”) to verify this claim. In Column 6, we show that the treatment groups benefited from preschools of much better quality in terms of equipment (+1.3 SD) and classroom setting (+0.4 SD). However, no treatment effects are observed for process quality and teacher characteristics between CPS and IPS.

times refereed to as robust lasso as it is robust to heteroskedasticity and clustering) selects almost the same variables except that it selects three baseline cognitive scores instead of one and select age instead of age squared.

⁴⁰The number of primary schools is not affected by the construction program as expected. Results not shown here.

⁴¹Note that some IPS nevertheless remain open in treatment group villages, either due to failure to implement a formal CPS or because they are run independently by a local pagoda or an NGO.

In order to investigate whether the CPS program improved preschools in domains likely to enhance child performance, we investigate the relationship between preschool quality and performance in the last panel of Table 4 (“Correlation quality-performance”). Here we analyze the correlation between each preschool quality index and midline cognitive skills, conditional on the baseline variables used for the reduced form estimations.⁴² We find that the process quality indices (“Pedagogy” and “Interactions”) are indeed significantly correlated with the midline cognitive development score while “Equipment”, the structural quality index that is the most affected by the construction program (+1.3 SD), shows no significant correlation with midline performance.

It is noteworthy that the indexes for “teacher characteristics and “Classroom setting” are also significantly correlated with the cognitive development index.⁴³ Overall, therefore, this analysis suggests that while the CPS program did improve some dimensions of quality, it may not have improved the domains that are the most associated with child cognitive development.

5.2 Preschool enrollment results

We now turn to child enrollment into preschool, a key stated objective for of the program.⁴⁴ Table 6 summarizes the impact on enrollment into each type of preschool, and into primary school, by treatment status. The measure of enrollment in this table is the child’s status on the day of the midline and endline survey visits. At midline, the interventions increased enrollment into any type of (pre)-school by 10.6 pp, compared to a control group enrollment rate of 43.5%.⁴⁵

The overall increase in preschool enrollment (observed on the day of the visit) is entirely driven by an increase in CPS enrollment of 41.0 pp at midline and 31.9 pp at endline (Column 2 of Table 6—the decline at endline is caused by the transition of six year-olds to primary school).⁴⁶ Differences in enrollment across treatment types are not statistically significant, with the exception of T2 villages, which have slightly higher CPS enrollment than T1 and T3 at endline (this may reflect the lower baseline availability of SPS in T2).

⁴²These regression coefficients can be interpreted as correlations since both dependant and independent variables are standardized. In addition, the R^2 is very high (65%). We note that this analysis is not causal.

⁴³This is somewhat surprising because the literature tends to suggest that observed teacher characteristics are not strongly correlated with students performance(Hanushek, 1971; Hanushek and Rivkin, 2006).

⁴⁴As mentioned above, follow-up data collection took place from April—June 2017 (midline) and May—July 2018 (endline); the school year for all types of preschools begins in late October or early November.

⁴⁵To interpret the results in a straightforward way, we do not include any control variables in this regression. Including them does not alter the results.

⁴⁶At endline, enrollment into formal CPS is not exactly 0 in the control group. A small number of children in a control group village in the province Kratie reported attending a formal CPS in an adjacent T3 village.

The increase in CPS enrollment is fueled in large part by a large substitution from IPS to CPS (IPS enrollment was 24.7 pp lower at midline and 20.2 pp lower at endline) which was by design as the new CPS replaced the IPS (Table 6). There was also a more moderate reduction in SPS enrollment (SPS enrollments were 5.6 pp lower at midline and 5.2 pp lower at endline). This reduction in SPS enrollment may affect how to interpret our estimates. First, since SPS are expected to outperform IPS, the substitution would have the effect of driving down the ITT estimates of the impact of the program on child development, particularly in T2. Second, as detailed in Section 4.3, the bounds exercise relies on the assumption that $LATE_{ac} \geq 0$. The reduction could be driven by two mechanisms: first, parents actually switching to SPS as a response to the program; second, the slight baseline imbalance in the availability of SPS (significant only in T2). Although we cannot empirically distinguish between these mechanisms, the consistency between the SPS availability and enrollment (Table 5 and Table 6) suggests that it is the imbalance that explains the result. In addition, the fact that SPS have higher quality than CPS makes it less likely that parents would switch. Both mechanisms would have the effect of driving down the ITT estimates of the impact of the program on child development, especially in T2.

Using the preschool enrollment data, we can calculate the share of *h-* and *a-compliers* (S_{ac} in Section 4.2). At midline, 52% of the sample are compliers, and 28% are *a-compliers*. Therefore, the share of *a-compliers* (among the compliers) is $S_{ac} = 54\% (=28/52)$ at midline. At endline, the total share of compliers goes down to 32% and is composed of a large majority of *a-compliers* (81%).

In addition to enrollment on the day of the survey visit, we analyze two other indicators of enrollment: ever enrolled and duration of enrollment, both measured by the time of the endline survey (Table 7). Note that at endline about 25% of the children are enrolled in primary school and primary school enrollment appears to have been unaffected by the CPS construction. The impacts on ever enrolled are consistent with those for the day of the visit. The cross-preschool-type patterns confirms that much of this increase comes from a large reduction in the probability that children are ever enrolled in an IPS and a more modest reduction in SPS enrollment.

These results suggest that about 28% of the eligible children did not attend any school at any point in time even when a preschool was available in the village. When asked why they did not send their child to preschool, parents indicate two main reasons. First, many parents expressed difficulties bringing and picking up the child at preschool (51%). Parents also often declared being afraid to let children go to school by themselves (62%) or that the school was too far away (35%). Since CPS sessions were only given in the morning for two hours, those parents who had to work outside of the village (in fields far from the village or in a garment factory) and/or

were living too far from the preschool could not send their child to school.⁴⁷ Second, many parents who did not enroll their child in preschool declare that the child was too “afraid”, not “ready” (“does not speak enough”) or not “mature” enough to go to school. Only a handful of parents declared that the preschool turned them down (8%) or that they could not afford to go school (< 1%). Only 2% of parents said that the school was too crowded or at capacity. These results suggest that supply-side constraints have only a limited bearing in explaining low enrollment. Similarly, very few parents invoked information constraints: only 3% declare that they “did not think about it.” Finally, in results not presented here, we find that T2 and T3 had no effect on the reasons given by parents for not sending their children to school.⁴⁸ This may in part explain why the T2 and T3 programs did not significantly increase demand for schooling in our context. We return to this issue in sub-section 5.4 when discussing the impact of demand-side and HBP interventions.

The intervention also significantly affected total exposure to (pre)school expressed in months. By endline, the intervention had increased the average months enrolled at (any) school by almost 1 month from a counterfactual of 7.4 months in the control group (Table 7). This is driven by an increase in the duration of CPS enrollment of 4.5 months, alongside a decrease in IPS (3.5 months) and SPS (0.58 months). The impact on total exposure at midline is similar to that at endline exposure (about one month). This is partly due to the enrollment in primary school at endline⁴⁹ and partly due to the substitution patterns.⁵⁰ This indicates that the endline results should not be interpreted as a two-years impact of the CPS program: it rather measures the one year effect on new younger preschool enrollees as well as the long-term effect on children enrolled in primary school and who benefited from preschool a year before. We discuss this point further when we present our results on child performance.

Finally, we assess heterogeneity in impacts on having ever enrolled in any type of school at endline (Table 8). Older children with higher baseline cognitive levels and from wealthier households tend to enroll more in CPS (significance across groups is indicated in bold in the table). For instance, children from the top wealth quartile are 10 pp more likely to enroll in CPS than children from the bottom wealth quartile. Similarly, children who performed in the top quartile at the baseline cognitive test are 12 pp more likely to attend CPS at endline. These results suggest possible

⁴⁷This explains why we do not report (and find) any impact on maternal work. The CPS actually represents a constraints to parents working in garment factories or in the field as they need to drop the children in the morning to school and pick them up two hours later.

⁴⁸Since the treatments affected enrollment, these analyses may not be causal but since T1, T2 and T3 experienced similar level of enrollment, their comparison remains approximately correct.

⁴⁹at endline, 26% of the sample (treatment and control) is already enrolled in primary school. These students therefore did not contribute to the treatment effect on exposure at endline

⁵⁰The children who enrolled between midline and endline were in large majority *a-compliers* Since the effect of exposure on *a-compliers* is zero, they contribute to reduce the treatment effect on exposure

positive complementarities between parental involvement and CPS enrollment. Note however that children from the top quartile of initial cognition and wealth are also more likely to substitute from SPS. The net effect of these opposite heterogeneous enrollment effects on follow-up developmental outcomes is uncertain.

5.3 Child development outcomes

5.3.1 One year impacts

ITT After one year the program significantly improved most domains of cognitive development (executive functions, numeracy, language and fine motor) in groups T1 and T3 (Table 9). Our index of cognitive development indicates that the cognitive performance of students in T1 and T3 improved by about 0.05 SD. The program also significantly reduced the occurrence of socio-emotional problems in all three treatment groups (by about 0.07 SD). Results are however not significantly different from zero in T2 for the cognitive domains of development. The lack of effect in T2 is puzzling because our baseline data did not show many differences between experimental groups (Table 3). In addition, our ITT estimates control for baseline test score, gender and age, which should account for any remaining differences between groups. The only noticeable difference between T2 and the control group is the lower availability of SPS in T2 (-11 pp at baseline and after, see Table 5). Subsequently to this initial imbalance, T2 children were significantly more likely to switch from alternative preschools (IPS and SPS) to CPS when the newly built preschool became available.⁵¹ Since SPS provide a much higher quality of education (see Table 4), this initial imbalance, that we attribute purely to chance, is likely to drive down the treatment effect in T2.

Sub-LATE The ITT results, as well as the $LATE_{cps}$ (reported in the first column of Table 11; +0.131 SD on the cognitive index), should be interpreted as a weighted average between the effect on students who would have attended an alternative preschool ($LATE_{ac}$) and the effect of students who would have stayed at home ($LATE_{hc}$), with expected larger impacts on the latter. We conduct our analysis of the subLATE parameters focusing only on groups C and T1. We do this for several reasons. First, since the group T2 has some imbalances in terms of preschool infrastructure, and since preschool infrastructure will be used to estimate the subLATEs, T2 is likely to affect the reliability of our subLATE estimates. Second, group T2 drives down the overall ITT effects, leaving little power to estimate subLATE parameters. Last, T2 and T3 include demand-side interventions which were supposed

⁵¹Indeed, lower availability of SPS at baseline is likely to have increased IPS enrollment and therefore generated a larger number of students switching from IPS to CPS in T2 once CPS became available (see Table 6. While these initial imbalances will not produce imbalances in terms of baseline child performance (as very few children were enrolled in SPS at baseline), they are likely to significantly reduce the treatment effect in T2

to increase overall CPS compliance but also to modify the socio-economic composition of the compliers. These demand-side interventions would therefore affect the validity of some of the assumptions on which the identification of the subLATEs relies.

Bounds We start our analysis of the subLATE parameters by bounding the $LATE_{hc}$ using the method developed in Section 4.3. This exercise first suggests that children who would have stayed at home absent the CPS construction spent between 3 and 10 months in (pre-)school (Table 10). For exposure, however, the $LATE_{hc}$ is probably closer to the upper bound. Indeed, children who switched from ASP to CPS are *unlikely* to have experienced a very different level of preschool exposure: we can assume that *a-compliers* spent about the same time at school irrespective of their treatment status, therefore the overall exposure to any school is likely to be about 10 months, one year after the baseline study. Our results on cognitive measures vary between 0.107 (low bounds for language) and 0.492 sd (upper bound for executive functions), many of them being significant or very close to significance level. The bounds for impact on the main index of cognitive development are 0.138 and 0.462 sd. Bounds for impacts on the socio-emotional index are 0.118 and 0.396 sd, but these are not statistically significant.

We then estimate the narrow bounds using the median of the ASP enrollment’s predicted value as the \mathbf{B} variable. Narrow bounds are as expected tighter (with the overall range between 0.087 and 0.354 sd)—with the magnitude of the upper bounds reduced and the lower bounds almost unaffected. The bounds for the index of cognitive development are 0.138 and 0.345 sd, consistent with moderate-to-large treatment effects. The effect on the socio-emotional index remain statistically insignificant.

Overall, these results confirm that low ITT impacts are consistent with moderate-to-large effects on h-compliers.

Conditional LATE We next turn to the conditional LATE approach which allows us to estimate both $LATE_{hc}$ and $LATE_{ac}$. In a first step we reject that each set of instruments is weak (the last row of Table 11). This is reassuring as it confirms that the instruments selected do predict ASP enrollment well in the control group (and its substitution to CPS in the treatment group). For each set of instruments, we also run an over-identification test (except for the LASSO which is just-identified): since the instruments are all from the baseline survey and are interacted with the random assignment, they are exogenous and will fail the over-identification test only when treatment effects are not homogeneous. Failing the over-identification test would therefore be an indication that the constant treatment assumption is violated. In most cases, our instruments pass the over-identification test. For both aggregate indexes, the over-identification test is passed for all specifications.

Using the instruments we selected (i.e. not those selected through LASSO), we find convincing evidence that (i) the $LATE_{hc}$ is larger than $LATE_{ac}$ in accordance with the main assumption used to identify the bounds, (ii) that the $LATE_{cps}$ is indeed a weighted average of both subLATEs, (iii) that the $LATE_{hc}$ on exposure to preschool is around nine months, again consistent with the results found with the bounds and (iv) that our preferred specification (province fixed effect, village and household characteristics) is very consistent with our bounding exercise, with $LATE_{hc}$ (+0.19 sd) estimated to be fairly close to the low bounds of Table 10.

The overall $LATE_{hc}$ for the cognitive index of development is driven by significant impacts on executive function, numeracy and language (around +0.2 sd each). It is dampened by an insignificant effect on fine motor skills.⁵² Consistently, the $LATE_{ac}$ is small and generally insignificant except for fine motor skills.⁵³

The results using LASSO are less precise, and impacts on neither index are statistically significant. However, point estimates for these indexes are in line with those from the other specifications however (particularly for the cognitive index).

Taken together, our SubLATE results suggest that although the extension of preschool availability improved outcomes for those who switched from home care, the children who had access to an alternative preschool did not benefit much from the new CPS. This is consistent with the fact that while the CPS had significantly better infrastructure (than IPS in particular), the construction program did not significantly improve teaching quality.

5.3.2 Endline results

At endline, the ITT effects are insignificant (Table 9). Impacts on all dimensions of child development are lower and we even detect a negative effect in T2 on fine motor skills (once again probably caused by initial imbalance). These differences are generally statistically significantly smaller (see “Midline vs Endline” panel). As mentioned above, the endline results should not be interpreted as the effect of staying two years in a preschool (compared to staying one year at midline). The treatment effect on overall preschool exposure does not actually increase between midline and endline and remains at about one month (see Table 7). This surprising result is easily explainable for two main reasons. First, many children who were enrolled in

⁵²Note that none of the $LATE_{hc}$ and $LATE_{ac}$ are significantly different from each other, except for exposure (in bold in Table 11)

⁵³The effect on fine motor skills is puzzling: if true, it violates the assumption of the bounds where *a-compliers* are assumed to benefit less from the new preschool than *h-compliers*, casting doubt on our bounding approach at least for this specific skill. Taken at face value, the fine motor result may be interpreted as a consequence of the improved CPS equipment compared to IPS, specifically in terms of chairs and tables. In that scenario, the IPS were so ill-equipped in terms of chairs and tables that they were doing poorly not only in comparison to the new CPS but also in comparison to what children could benefit from at home, as a $LATE_{ac} > LATE_{hc}$ implies that IPS did worse than home. This interpretation is consistent with Table 4 where we found that CPS essentially improved the quality of the infrastructure and not the process quality.

a CPS at midline are enrolled in a primary school at endline and are therefore not increasing their total preschool exposure. Second, a large number of children enrolled in a CPS at midline but not in primary school at endline would have attended an alternative preschool absent the construction and are therefore not contributing to increasing overall exposure. As a result, while CPS exposure does increase slightly at endline (+4.5 months compared to +3.6 months at midline), it is in large part compensated by alternative preschool substitution as shown in Table 7.⁵⁴

This pattern of enrollment may explain in part the absence of result at endline. At endline, only 6.5% of the sample is *h-compliers* (i.e children who would have stayed at home otherwise) and therefore most of the new CPS beneficiaries would have attended other preschools. Since our subLATE analysis suggests that the $LATE_{ac}$ is close to zero⁵⁵, this high degree of substitution is likely to drive the effect toward zero. Finally, the absence of impacts at endline suggests that the treatment children who were exposed to more preschool before enrolling in primary school, and who benefited from a small treatment effect at midline, did not perform significantly better than the control group in primary school: it appears that the control group quickly caught up with the treatment group once both groups were enrolled in primary school⁵⁶.

5.3.3 Heterogeneous effects

These overall results may mask heterogeneous impacts. At midline, we find that it is non-stunted children who are driving the effects, suggesting that the CPS were unable to foster development of the most-disadvantaged children (Table 12). This is consistent with the fact that the program only significantly impacted children whose household education level was in quartiles 3 and 4.

A similar pattern appears at endline where the absence of results on the whole sample masks heterogeneous effects: Children from the top baseline ability quartile, from the highest wealth quartile and living in households with higher level of education show slightly larger impacts than the rest of the sample (in bold in Table 12). While these effects are not always significantly different from zero, they are close to significance and display less reduction as compared to midline impacts than for other groups.⁵⁷ Given our low level of precision in sub-group analysis, these heterogeneity results should be taken with a grain of salt. They would for instance

⁵⁴overall only 12% of the children stayed more than one year in (any) preschool.

⁵⁵The absence of ITT effect makes it impossible to identify the subLATE parameter at endline. Our subLATE at endline are therefore not differentiable from zero.

⁵⁶In an analysis not shown here, we show that primary school children do not fare better in the treatment group than in the control group at endline. Since this analysis is not fully identified (since the program could have had an effect on primary school enrollment and composition), we do not present the result of this analysis here.

⁵⁷Again, these ITT impacts potentially is likely to mask larger impacts on the children who would have stayed at home absent the treatment, yet our capacity to detect subLATE is low at endline. we therefore do not run a full subLATE analysis at endline.

not pass a correction for multi-hypothesis testing.

Yet, if taken at face value, these heterogeneous impacts could be explained by differences in preschool enrollment since children of the wealthiest or most educated quartiles enroll more in CPS (Table 8). Yet, only children from the fourth quartile (of initial cognitive ability, household education, and wealth) show positive cognitive impacts while differential enrollment were found for children in quartiles 2 and 3 as well. In addition, children from the fourth quartile are significantly more affected by the SPS substitution⁵⁸ (and to a lesser extent to a IPS substitution) which should, in theory, drive down their treatment effect since we established that the $LATE_{ac}$ is close to zero. Differential take-up rates cannot entirely account for these heterogeneous effects. Alternatively, we can speculate that these heterogeneous effects suggest that either (i) the preschool content was more adjusted children belonging to the top quartiles (of education, wealth or initial cognition), (ii) these children benefited from a more favorable environment at home that complemented the skilled learned at school (e.g. books, toys, better health, electric light...), or (iii) preschool enrollment (indirectly) caused parenting to increase or improve. We study the latter by looking at the quality of parenting at midline and endline (Table B6) and find that the program positively affected cognitive parenting at midline and to a lesser extent at endline.⁵⁹ This effect is stronger on children living in wealthier or more educated families at midline (Table B7).⁶⁰ These heterogeneous parental impacts are barely significant and disappear at endline however, suggesting that they are unlikely to be the only mechanism since the effect on their children's development are sustained at endline.

Overall, our heterogeneous findings are not very significant and should therefore be taken with caution. Yet, our results suggest that relatively more privileged children—children with wealthier or more educated parents— benefited more from the CPS. These heterogeneous effects cannot be entirely explained by larger levels of preschool enrollment among these sub-groups of children neither by larger involvement of their parents. More research needs to be conducted to better understand the possible complementarity with early childcare programs.

⁵⁸As mentioned, we have reasons to believe that this substitution is only apparent and is driven by initial imbalance in SPS availability

⁵⁹Table B6 further documents that there is no substitution between parental involvement and preschool availability.

⁶⁰A similar pattern is observed by Padilla (2019) who documents effects of the Head Start program in the US on parenting outcomes. She finds positive effects on parents' cognitively stimulating behaviors and no effects on parent's use of harsh punishment. The effects on cognitively stimulating behaviors are driven by parents with higher cognitive stimulation prior the intervention.

5.4 D2D and HBP interventions

We finally turn to the impacts of D2D and HBP interventions on enrollment, starting with an assessment of “take up” (Table 13).⁶¹ For D2D, no treatment groups report any additional home visits as a result of the program. It is possible that home visits from the village head are difficult to capture (and for respondents to recall) due to their informal nature. It is also possible that village heads did not increase the number of home visits (perhaps because they had no incentive to do so). Both interpretations are consistent with the fact that almost 70% of respondents in control villages report having received a home visit.

Respondents in T2 and T3 were more likely to report having received an information leaflet whereas those in T1 were not (as one would expect given the content of the different treatments). This impact is apparent at both midline (+ 5.6 pp for T2 and +7.9 pp for T3) and endline (+11 pp for T2 and +8.3 pp for T3). While the result is somewhat reassuring—coverage varies significantly across treatment arms—the overall percentage of respondents who report receiving a leaflet is small in all groups. The highest take-up by endline is reported in T2 villages, where it only reaches 18%. Even allowing for poor respondent recall, and despite the field team’s effort to ensure that 100% of the T2 and T3 sample received a leaflet⁶², these results are fairly disappointing and, together with the absence of additional home visits, makes the D2D treatment particularly weak.

The patterns in the likelihood of having participated in an HBP (either “ever” or “more than once”) are somewhat mixed. Almost all T3 villages eventually benefited from an HBP (89%, Table 5). The impact of having been assigned to T3 significantly increases the likelihood of having participated in an HBP (+19 pp at midline and +10 pp at endline). But the impact of being assigned to T1 is also statistically significant (although about half the magnitude). This is surprising since these villages had no formal HBP deployed as a part of the intervention. At the same time, and consistent with the discussion above, the overall take-up rates implied by these numbers are small. The highest participation rate reported for ever having participated in an HBP by endline is just below 30% (T3), again suggesting the intensity of the intervention was weak. The lack of intensity of HBP, due to low caregiver participation, is perhaps not surprising though, as this program had already been shown to have little impact overall (Bouguen et al., 2018).

Despite these statistically significant (albeit small) differences in the intensity of the D2D and HBP interventions across groups, we find that they did not significantly increase enrollment (Table 13), they did not modify the performance of the CPS

⁶¹Self-reported exposure to the D2D (home visit + leaflet) and HBP interventions is likely to be less reliable than for preschool enrollment since these were often just a one-time interaction many months in the past. Nevertheless,

⁶²Recall that the field team directly conducted the distribution of leaflet, as well as its explanation. We are confident that close to 100% of the respondents receive the leaflet. We interpret this take-up measure as how salient the information included in the leaflet was for the caregiver.

(Table 9), and they did not affect parenting behaviors (Appendix Tables B6 and B7).

Neither the program as a whole, nor the D2D and HBP interventions, seems to have affected the way parents perceived education. While the program did have an effect on the reported optimal age for starting preschool, these impacts are not larger in T2 and T3 (Table B8). In addition, there are no impacts on the perceived return to school, maybe because parents already considered the return to primary and secondary school to be high (6.8% per school year in primary and 8.5% for secondary in the control group) and even higher than those reported by Humphreys (2015) based on Cambodia’s Socio-Economic Survey (CSES) in 2010. Unlike other contexts (for example in Madagascar (Nguyen, 2008) or in Dominican Republic (Jensen, 2010)), Cambodian caregivers did not appear to have under-estimated the return to education. We cannot rule out the possibility that a more intensive set of programs (e.g., more frequent home visits, more intensive HBP) might have had bigger impacts. But it is also possible that, in the context of these small rural villages, the construction of the CPS itself had a demand-side “information” component in the sense that it was likely an important village event that households would have known about. In such a context, it could have been the case that the additional demand-side programs did not convey any information over and above the construction itself given what parents already believed about the return to education.

6 Conclusion

Consistent with findings in other studies, our results show that a relatively well-implemented preschool program—90% of treatment villages received a community preschool (CPS)—can initially increase the cognitive and socio-emotional performance of young children. Even though the midline ITT impacts are low in magnitude (+0.04 SD), the complex substitution patterns make these results consistent with moderate-to-large impacts on children who would have stayed at home absent the program. Our analysis also shows that the impacts on children who would have attended another preschool absent the CPS construction are small and insignificant at midline. These suggest that the midline ITT impacts are driven by additional enrollments rather than improvement in the quality of the available preschool provision. This is consistent with our findings related to preschool quality: While the CPS program considerably improved the quality of infrastructure, we did not find significant difference between CPS and the informal preschool program in terms of teaching quality.

Our results also show no significant overall treatment effect at endline. The absence of effect at endline can in part be explained by the fact that overall exposure to preschool remains equivalent at midline and endline. This is due to many children

transitioning to primary school and children having access to alternative preschools in the control group. The absence of effect at endline nevertheless suggests that the midline benefits from the CPS did not persist once children entered primary school.

Last, our average results mask variation across subgroups. The relatively wealthier and more educated households whose children performed well at baseline are performing slightly better at midline, as are children who have higher cognitive skills to begin with. At endline, the impact of the preschool program is persistent only for these categories of children. Although these results are barely significant and should therefore be taken with caution, they suggest that either the CPS program was better adjusted to children already more cognitively developed, or that there exists positive complementarity between the home environment and the education provided in the preschools. At midline, we have some evidence that the quality of parenting improved among the wealthier children. However, this effect disappears at endline suggesting that parental involvement is unlikely to be the only mechanism. The lack of effects on the least-privileged children might be a consequence of the short duration of the CPS shift (2 hours). By reducing the influence of the home environment, longer shifts could perhaps have had better outcomes for these children. Further research would need to be conducted to establish the optimal intensity of such program and the possible complementarity between preschool and home environment.

While the intervention we study succeeded in providing better infrastructure and materials relative to informal arrangements, it did not provide a substantively better teaching and learning experience. The results therefore suggest that in Cambodia, and likely in other similar contexts, a key focus should be on improving the process quality aspects of preschool, namely improvements in teacher pedagogical skills and ultimately teacher-child interactions. This might require better training of teachers but also perhaps approaches that build on intrinsic and potentially extrinsic motivation of teachers to ensure that they are supported in, and recognized for, putting better teaching methods into practice.

Our results finally show that the demand-side interventions implemented were unable to boost the participation in and performance of the CPS. These approaches were not enough to mobilize much additional enrollment over and above simply building preschools, suggesting that other factors likely drive the decision to send children to preschool. In Cambodia, households have a high perceived return to education and are generally willing to send their children to schools. Other constraints related to the practicality of the preschool shift (only 2 hours in the morning while families work in the field or garment factory) or the distance to the preschool appeared to be much more relevant to caregivers in our context. Likewise, while HBP were established on time (89%), they failed to attract a sufficient number of families to have any visible impacts on the enrollment in CPS or the performance of children.

Our findings suggest that more work is required to understand the drivers of preschool demand and quality. Direct or indirect costs may play a role, so approaches to reduce those costs—for example, cash transfers or even further reductions in travel distances—might be necessary to induce higher participation rates. In addition, if the quantity and quality of preschool services do not meet families’ needs, households might have low demand for them. Therefore, increasing the quantity (time spent in preschool per day) or the quality of the preschool offer may increase demand. Moreover, the fact that children from wealthier families did benefit from the newly constructed preschools suggests that other demand-side factors (such as complementary investments in child development) act as additional constraints. Further investigation of these factors is required to better understand and address them.

References

- Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature* 56(4), 1360–1446.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association* 103(484), 1481–1495.
- Andrew, A., O. Attanasio, R. Bernal, L. C. Sosa, S. Krutikova, and M. Rubio-Codina (2019, August). Preschool quality and child development. Working Paper 26191, National Bureau of Economic Research.
- Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics* 131(3), 1415–1453.
- Araujo, M. C., M. Dornal, and N. Schady (2019). Childcare quality and child development. *Journal of Human Resources* 54(3), 656–682.
- Baker, M., J. Gruber, and K. Milligan (2008, August). Universal Child Care, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy* 116(4), 709–745.
- Baker, M., J. Gruber, and K. Milligan (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy* 11(3), 1–26.
- Belfield, C. R., M. Nores, S. Barnett, and L. Schweinhart (2006). The high/scope perry preschool program cost–benefit analysis using data from the age-40 followup. *Journal of Human resources* 41(1), 162–190.
- Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6), 2369–2429.
- Belloni, A., V. Chernozhukov, and C. Hansen (2013, 11). Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies* 81(2), 608–650.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014, May). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28(2), 29–50.
- Berkes, J., A. Bouguen, D. Filmer, and T. Fukao (2019). Combining supply and demand-side interventions : Evidence from a large preschool program in cambodia : Impact evaluation final report. <http://documents.worldbank.org/curated/en/110771561664132545/Impact-Evaluation-Final-Report>.
- Berkes, J., A. Raikes, A. Bouguen, and D. Filmer (2019). Joint roles of parenting and nutritional status for child development: Evidence from rural cambodia. *Developmental Science* 22(5), e12874.
- Berlinski, S., S. Galiani, and P. Gertler (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics* 93(1), 219–234.
- Berlinski, S., S. Galiani, and M. Manacorda (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics* 92(5), 1416–1440.

- Bernal, R., O. Attanasio, X. Peña, and M. Vera-Hernández (2019). The effects of the transition from home-based childcare to childcare centers on children’s health and development in colombia. *Early Childhood Research Quarterly* 47, 418–431.
- Black, M. M., S. P. Walker, L. C. Fernald, C. T. Andersen, A. M. DiGirolamo, C. Lu, D. C. McCoy, G. Fink, Y. R. Shawar, J. Shiffman, et al. (2017). Early childhood development coming of age: Science through the life course. *The Lancet* 389(10064), 77–90.
- Blimpo, M. P., P. M. Amaro Da Costa Luz Carneiro, P. Jervis, and T. Pugatch (2019). Improving access and quality in early childhood development programs: Experimental evidence from the gambia.
- Blimpo, M. P. and T. Pugatch (2017). Scaling up children’s school readiness in the gambia: Lessons from an experimental study. *Working paper*.
- Bouguen, A., D. Filmer, K. Macours, and S. Naudeau (2013). Impact evaluation of three types of early childhood development interventions in cambodia (english). *Policy Research working paper IE 97*(WPS 6540).
- Bouguen, A., D. Filmer, K. Macours, and S. Naudeau (2018). Preschool and parental response in a second best world: Evidence from a school construction experiment. *Journal of Human Resources* 53(2), 474–512.
- Brinkman, S. A., A. Hasan, H. Jung, A. Kinnell, and M. Pradhan (2017). The impact of expanding access to early childhood education services in rural indonesia. *Journal of Labor Economics* 35(S1), S305–S335.
- Britto, P. R., S. J. Lye, K. Proulx, A. K. Yousafzai, S. G. Matthews, T. Vaivada, R. Perez-Escamilla, N. Rao, P. Ip, L. C. Fernald, et al. (2017). Nurturing care: Promoting early childhood development. *The Lancet* 389(10064), 91–102.
- Britto, P. R., H. Yoshikawa, and K. Boller (2011). Quality of early childhood development programs in global contexts: Rationale for investment, conceptual framework and implications for equity. social policy report. volume 25, number 2. *Society for Research in Child Development*.
- Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics* 1(4), 200–232.
- Carneiro, P. and R. Ginja (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from head start. *American Economic Journal: Economic Policy* 6(4), 135–73.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1), C1–C68.
- Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy* 126(6), 2356–2409.
- Cunha, F. and J. Heckman (2007). The technology of skill formation. *The American Economic Review* 97(2), 31.
- Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives* 27(2), 109–132.
- Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015, November). *Early*

- Childhood Education*, pp. 235–297. University of Chicago Press.
- Engle, P. L., L. C. Fernald, H. Alderman, J. 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), 1339–1353.
- Feller, A., T. Grindal, L. Miratrix, and L. C. Page (2016, 09). Compared to what? variation in the impacts of early childhood education by alternative care type. *Ann. Appl. Stat.* 10(3), 1245–1285.
- Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science* 344(6187), 998–1001.
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review* 61(2), 280–288.
- Hanushek, E. A. and S. G. Rivkin (2006). Teacher quality. *Handbook of the Economics of Education* 2, 1051–1078.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–86.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science* 312(5782), 1900–1902.
- Henderson, R. H. and T. Sundaresan (1982). Cluster sampling to assess immunization coverage: A review of experience with a simplified sampling method. *Bulletin of the World Health Organization* 60(2), 253.
- Hull, P. (2018). Isolateing: Identifying counterfactual-specific treatment effects with cross-stratum comparisons. *Working Paper*.
- Humphreys, J. (2015). Education premiums in cambodia: Dummy variables revisited and recent data. *Econ Journal Watch* 12(3), 339–345.
- Ichino, A., M. Fort, and G. Zanella (2019). Cognitive and non-cognitive costs of daycare 0-2 for children in advantaged families. *Journal of Political Economy*.
- Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125(2), 515–548.
- Johnson, R. C. and C. K. Jackson (2018, February). Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. NBER Working Paper Series 23489, National Bureau of Economic Research.
- Kline, P. and C. R. Walters (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics* 131(4), 1795–1848.
- Lee, D. S. (2009, 07). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76(3), 1071–1102.
- List, J. A., D. Suskind, and L. H. Supplee (2021). *The Scale-up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and what We Can Do about it*. Routledge.

- Martinez, S., S. Naudeau, and V. Pereira (2017a). The promise of preschool in africa: A randomized impact evaluation of early childhood development in rural mozambique. *Washington, DC: The World Bank*.
- Martinez, S., S. Naudeau, and V. A. Pereira (2017b). Preschool and child development under extreme poverty: Evidence from a randomized experiment in rural mozambique. *World Bank Policy Research Working Paper* (8290).
- MoEYS (2014). Education strategic plan 2014-2018. Technical report, Kingdom of Cambodia, Ministry of Education, Youth and Sport.
- MoEYS (2017). The education, youth and sport performance in the academic year 2015-2016 and goals for the academic year 2016-2017. Technical report, Kingdom of Cambodia, Ministry of Education, Youth and Sport.
- Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript 6*.
- Padilla, C. M. L. (2019). *Moving Beyond the Average Effect: Quantifying and Exploring Variation in Head Start Treatment Effects on Parenting Behavior*. Ph. D. thesis, Georgetown University.
- Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer (2012). Third grade follow-up to the head start impact study. Technical report, Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Raikes, A., N. Koziol, M. Janus, L. Platas, T. Weatherholt, A. Smeby, and R. Sayre (2019). Examination of school readiness constructs in tanzania: Psychometric evaluation of the melqo scales. *Journal of Applied Developmental Psychology* 62, 122–134.
- Rao, N., J. Sun, J. M. Wong, B. Weekes, P. Ip, S. Shaeffer, M. Young, M. Bray, E. Chen, and D. Lee (2014). *Early Childhood Development and Cognitive Development in Developing Countries*. DFID, UK Government.
- Shonkoff, J., J. Richmond, P. Levitt, S. Bunge, J. Cameron, G. Duncan, and C. Nelson III (2016). From best practices to breakthrough impacts a science-based approach to building a more promising future for young children and families. *Cambirdge, MA: Harvard University, Center on the Developing Child*.
- UNESCO, UNICEF, B. (2017). Overview melqo: Measuring early learning quality outcomes.
- UNICEF (2019). *A World Ready to Learn: Prioritizing Quality Early Childhood Education*. UNICEF, New York, April 2019.
- World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise*. The World Bank.
- Zelazo, P. D. (2006). The dimensional change card sort (dccs): A method of assessing executive function in children. *Nature Protocols* 1(1), 297.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101(476), 1418–1429.

Table 1: Random Treatment Allocation

| Group | CPS | D2D | HBP | Villages |
|---------|-------------------------------------|-------------------------------------|-------------------------------------|----------|
| T1 | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 120 |
| T2 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | 64 |
| T3 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 63 |
| Control | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | 58 |

The Table shows the treatment allocation of the 305 villages in the three treatment groups and the control group.

Table 2: CPS Construction Timetable

| Period | Activity | CPS construction |
|-------------------|--------------------------|------------------|
| 03/2016 | Begin CPS construction | 0% completed |
| 05/2016 – 07/2016 | Baseline data collection | 17% completed |
| 10/2016 | Beginning of school year | 82% completed |
| 12/2016 | BY end of 2016 | 86% completed |
| 04/2017 – 06/2017 | Midline data collection | 89% completed |
| 05/2018 – 07/2018 | Endline data collection | 91% completed |

The Table shows the construction timetable of the CPS.

Table 3: Balancing

| | N | C | T1 | T2 | T3 | Any T |
|--|--------|-------|---------------------|---------------------|---------------------|---------------------|
| Child characteristics | | | | | | |
| Cognitive index | 7,491 | 0.000 | -0.001 (0.034) | -0.002 (0.040) | -0.007 (0.038) | -0.003 (0.032) |
| Socioemotional problems | 7,472 | 0.000 | -0.027 (0.042) | -0.064 (0.045) | -0.046 (0.047) | -0.041 (0.039) |
| Age (yrs) w. decimals | 8,589 | 3.409 | -0.021 (0.027) | 0.027 (0.030) | -0.047 (0.031) | -0.016 (0.024) |
| Female | 34,807 | 0.542 | -0.003 (0.006) | -0.013** (0.007) | -0.001 (0.007) | -0.005 (0.005) |
| Stunted (lhfa<2sd) | 7,473 | 0.341 | 0.018 (0.018) | 0.020 (0.022) | 0.032 (0.022) | 0.022 (0.017) |
| Household and caregiver characteristics | | | | | | |
| Household size | 41,754 | 6.568 | -0.014 (0.122) | 0.076 (0.153) | 0.016 (0.134) | 0.017 (0.114) |
| Wealth score | 34,803 | 0.000 | 0.025 (0.071) | 0.064 (0.081) | 0.131* (0.073) | 0.063 (0.064) |
| Caregiver female | 41,754 | 0.461 | 0.000 (0.032) | -0.048 (0.040) | -0.009 (0.036) | 0.010 (0.021) |
| Caregiver age | 7,634 | 40.77 | 0.425 (0.642) | -1.180 (0.741) | 0.179 (0.679) | 0.023** (0.011) |
| Caregiver years of educ. | 7,626 | 4.683 | -0.290 (0.428) | -0.266 (0.450) | 0.291 (0.495) | -0.130 (0.392) |
| Caregiver Raven score | 7,562 | 0.051 | -0.072 (0.045) | 0.009 (0.048) | -0.007 (0.049) | -0.034 (0.041) |
| Cognitive parenting | 7,615 | 0.000 | 0.023 (0.042) | -0.003 (0.050) | 0.080* (0.044) | 0.031 (0.038) |
| Socioemotional parenting | 7,616 | 0.000 | -0.012 (0.037) | -0.118** (0.047) | -0.020 (0.044) | -0.041 (0.035) |
| Negative parenting | 7,617 | 0.000 | 0.047 (0.043) | 0.087* (0.047) | 0.013 (0.044) | 0.048 (0.039) |
| Baseline program attendance | | | | | | |
| Attending preschool | 7,616 | 0.153 | 0.073*** (0.023) | 0.063** (0.027) | 0.108*** (0.026) | 0.080*** (0.021) |
| Participated in HBP | 7,613 | 0.104 | 0.031* (0.017) | 0.056*** (0.020) | 0.184*** (0.023) | 0.078*** (0.017) |
| Received D2D | 7,613 | 0.017 | 0.003 (0.006) | 0.012* (0.007) | 0.021*** (0.007) | 0.010* (0.006) |

Differences in means are based on OLS regressions at the child level on province fixed effects and three treatment group dummies (columns 3-5) or a joint treatment group dummy (column 6). Robust standard errors clustered at village level in parentheses. Control group standard deviations in square brackets.

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Structural and process quality by type of preschool

| | Preschool comparison | | | | | Experimental comparison | | Correlation quality-performance | | |
|---------------------------|----------------------|------------|------------|----------------------|----------------------|-------------------------|-------------|---------------------------------|-------------|---------------|
| | CPS (1) | IPS (2) | SPS (3) | IPS-CPS (4) | SPS-CPS (5) | Any T-C (6) | Obs. (7) | Cognitive index (8) | Obs. (9) | R^2 (10) |
| Structural quality | | | | | | | | | | |
| Teacher char. | -0.34 | -0.31 | 1.40 | 0.029 (0.106) | 1.743*** (0.121) | -0.003 (0.165) | 326 | 0.082*** (0.017) | 2697 | 0.623 |
| Classroom setting | 0.08 | -0.51 | 0.22 | -0.592*** (0.148) | 0.138 (0.136) | 0.395*** (0.141) | 326 | 0.087*** (0.016) | 2706 | 0.625 |
| Equipment | 0.35 | -1.07 | -0.18 | -1.419*** (0.157) | -0.539*** (0.125) | 1.306*** (0.163) | 325 | 0.016 (0.017) | 2702 | 0.619 |
| Process quality | | | | | | | | | | |
| Pedagogy | -0.05 | -0.23 | 0.38 | -0.179 (0.141) | 0.428*** (0.152) | 0.120 (0.148) | 327 | 0.044** (0.017) | 2709 | 0.622 |
| Interactions | -0.06 | -0.10 | 0.29 | -0.044 (0.139) | 0.347** (0.136) | -0.017 (0.151) | 313 | 0.033* (0.017) | 2606 | 0.620 |

Columns under “Preschool comparison” show averages by type of preschool and their comparison. Columns under “Experimental comparison” show the difference between the treatment and control preschools. Columns under “Correlation quality-performance” give the results of the regression of the Midline cognitive development index on the preschool quality measures, controlling for usual baseline variables. We use robust standard errors, clustered at village level for the quality-performance analysis. For details about individual variables of summary scores see Table B2 and B3. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Preschool infrastructure

| | Obs. | C | T1 | T2 | T3 | Any T |
|-----------------|------|------------------|----------------------|----------------------|----------------------|----------------------|
| Baseline | | | | | | |
| Any preschool | 305 | 0.759 [0.432] | 0.008 (0.069) | 0.003 (0.078) | 0.085 (0.073) | 0.027 (0.062) |
| CPS | 305 | 0.017 [0.131] | 0.199*** (0.042) | 0.268*** (0.060) | 0.248*** (0.058) | 0.230*** (0.032) |
| IPS | 305 | 0.603 [0.493] | -0.128 (0.079) | -0.143 (0.090) | -0.057 (0.090) | -0.114 (0.072) |
| SPS | 305 | 0.224 [0.421] | -0.057 (0.065) | -0.145** (0.065) | -0.084 (0.070) | -0.086 (0.059) |
| HBP | 305 | 0.397 [0.493] | 0.217*** (0.079) | 0.254*** (0.089) | 0.400*** (0.082) | 0.274*** (0.071) |
| # of preschools | 305 | 1.159 [0.428] | -0.029 (0.073) | -0.076 (0.076) | 0.008 (0.082) | -0.030 (0.068) |
| Midline | | | | | | |
| Any preschool | 305 | 0.810 [0.395] | 0.123** (0.057) | 0.190*** (0.052) | 0.174*** (0.054) | 0.153*** (0.053) |
| CPS | 305 | 0.000 [0.000] | 0.858*** (0.032) | 0.984*** (0.016) | 0.937*** (0.030) | 0.911*** (0.018) |
| IPS | 305 | 0.655 [0.479] | -0.564*** (0.068) | -0.623*** (0.067) | -0.546*** (0.074) | -0.574*** (0.065) |
| SPS | 305 | 0.241 [0.432] | -0.058 (0.067) | -0.130* (0.069) | -0.070 (0.074) | -0.079 (0.061) |
| HBP | 305 | 0.362 [0.485] | 0.105 (0.078) | 0.162* (0.090) | 0.591*** (0.069) | 0.245*** (0.071) |
| # of preschools | 305 | 0.948 [0.575] | 0.177** (0.089) | 0.179** (0.086) | 0.286*** (0.098) | 0.206** (0.081) |
| Endline | | | | | | |
| Any preschool | 305 | 0.845 [0.365] | 0.097* (0.052) | 0.155*** (0.048) | 0.140*** (0.050) | 0.123** (0.049) |
| CPS | 305 | 0.017 [0.131] | 0.841*** (0.036) | 0.983*** (0.017) | 0.905*** (0.038) | 0.894*** (0.025) |
| IPS | 305 | 0.724 [0.451] | -0.607*** (0.066) | -0.677*** (0.065) | -0.599*** (0.072) | -0.623*** (0.062) |
| SPS | 305 | 0.259 [0.442] | -0.050 (0.069) | -0.132* (0.072) | -0.040 (0.078) | -0.068 (0.063) |
| HBP | 305 | 0.259 [0.442] | 0.191*** (0.074) | 0.281*** (0.086) | 0.632*** (0.070) | 0.328*** (0.066) |
| # of preschools | 305 | 0.983 [0.577] | 0.192** (0.091) | 0.211** (0.097) | 0.271*** (0.101) | 0.217*** (0.083) |

Table shows preschools available in sample villages. C is the control group mean. Columns 3-6 show the difference between treatment group and control group mean based on a regression of preschool availability on a set of dummy variables for each treatment group (columns 3-5) or on a dummy variable for all treatment groups (column 6). No control variables included in regression model. Estimates correct for heteroskedasticity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Enrollment on day of survey by type of school

| | Obs | C | T1 | T2 | T3 | Any T |
|--------------------------------|-------|------------------|----------------------|----------------------|----------------------|----------------------|
| Midline (3-5 years old) | | | | | | |
| Any school | 6,992 | 0.435 [0.496] | 0.106*** (0.036) | 0.095** (0.041) | 0.119*** (0.039) | 0.106*** (0.033) |
| CPS | 6,992 | 0.000 [0.000] | 0.389*** (0.025) | 0.432*** (0.027) | 0.426*** (0.030) | 0.410*** (0.016) |
| ASP | 6,992 | 0.435 [0.496] | -0.283*** (0.034) | -0.337*** (0.032) | -0.307*** (0.036) | -0.303*** (0.032) |
| ... IPS | 6,992 | 0.284 [0.451] | -0.247*** (0.034) | -0.263*** (0.033) | -0.232*** (0.037) | -0.247*** (0.033) |
| ... SPS | 6,992 | 0.112 [0.315] | -0.041* (0.024) | -0.075*** (0.023) | -0.066*** (0.023) | -0.056** (0.022) |
| ... Primary school | 6,992 | 0.040 [0.196] | 0.004 (0.009) | 0.001 (0.010) | -0.010 (0.009) | 0.000 (0.008) |
| Endline (4-6 years old) | | | | | | |
| Any school | 7,015 | 0.658 [0.474] | 0.066** (0.029) | 0.055* (0.033) | 0.060* (0.033) | 0.062** (0.027) |
| CPS | 7,015 | 0.009 [0.093] | 0.297*** (0.021) | 0.373*** (0.027) | 0.307*** (0.028) | 0.319*** (0.015) |
| ASP | 7,015 | 0.650 [0.477] | -0.231*** (0.031) | -0.317*** (0.033) | -0.247*** (0.033) | -0.257*** (0.028) |
| ... IPS | 7,015 | 0.258 [0.437] | -0.194*** (0.031) | -0.228*** (0.031) | -0.191*** (0.034) | -0.202*** (0.030) |
| ... SPS | 7,015 | 0.135 [0.342] | -0.041 (0.027) | -0.066** (0.027) | -0.058** (0.027) | -0.052** (0.025) |
| ... Primary school | 7,015 | 0.257 [0.437] | 0.004 (0.024) | -0.024 (0.025) | 0.002 (0.026) | -0.004 (0.022) |

The table provides the ITT treatment effects on preschool enrollment measured at the day of the midline and endline survey. T1, T2, T3 and any T give the treatment effects. No additional control variables included in regression model. C mean gives the average in the control group and the bottom rows provide the p-value of the comparison between the treatment coefficients. Estimates correct for heteroskedasticity and within-village correlations.

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Ever enrolled and enrollment exposure at endline by type of school

| | Obs | C | T1 | T2 | T3 | Any T |
|--------------------------------------|-------|------------------|----------------------|----------------------|----------------------|----------------------|
| Ever enrolled at endline | | | | | | |
| Any school | 7,015 | 0.718 [0.450] | 0.061** (0.029) | 0.064** (0.032) | 0.079** (0.031) | 0.067** (0.027) |
| Preschool | 7,015 | 0.591 [0.492] | 0.094** (0.039) | 0.112*** (0.043) | 0.126*** (0.041) | 0.107*** (0.036) |
| CPS | 7,015 | 0.017 [0.129] | 0.478*** (0.028) | 0.595*** (0.029) | 0.528*** (0.036) | 0.521*** (0.019) |
| ASP | 7,015 | 0.708 [0.455] | -0.254*** (0.031) | -0.355*** (0.034) | -0.268*** (0.034) | -0.283*** (0.028) |
| ... IPS | 7,015 | 0.413 [0.493] | -0.301*** (0.040) | -0.357*** (0.038) | -0.286*** (0.044) | -0.311*** (0.038) |
| ... SPS | 7,015 | 0.200 [0.400] | -0.057 (0.036) | -0.106*** (0.035) | -0.088** (0.035) | -0.078** (0.032) |
| ... Primary school | 7,015 | 0.262 [0.440] | 0.004 (0.024) | -0.023 (0.025) | 0.002 (0.026) | -0.003 (0.022) |
| Exposure in months at midline | | | | | | |
| Preschool | 6,985 | 3.405 [5.513] | 0.907** (0.404) | 0.797* (0.457) | 1.386*** (0.444) | 1.006*** (0.362) |
| CPS | 6,985 | 0.000 [0.000] | 3.421*** (0.253) | 3.651*** (0.297) | 3.872*** (0.321) | 3.599*** (0.167) |
| ASP | 6,985 | 3.672 [5.627] | -2.411*** (0.346) | -2.783*** (0.344) | -2.492*** (0.365) | -2.527*** (0.333) |
| ... IPS | 6,985 | 2.411 [4.858] | -2.088*** (0.320) | -2.250*** (0.317) | -1.945*** (0.347) | -2.092*** (0.318) |
| ... SPS | 6,985 | 0.994 [3.404] | -0.426* (0.219) | -0.604*** (0.220) | -0.541** (0.221) | -0.502** (0.206) |
| Exposure in months at endline | | | | | | |
| Any school | 7,017 | 7.423 [7.180] | 0.914* (0.481) | 0.752 (0.537) | 1.209** (0.511) | 0.951** (0.440) |
| Preschool | 7,017 | 5.370 [6.160] | 0.837* (0.451) | 0.934* (0.503) | 1.161** (0.481) | 0.946** (0.414) |
| CPS | 7,017 | 0.102 [0.892] | 4.129*** (0.268) | 5.088*** (0.294) | 4.530*** (0.354) | 4.476*** (0.182) |
| ASP | 7,017 | 7.322 [7.164] | -3.215*** (0.458) | -4.336*** (0.453) | -3.320*** (0.473) | -3.525*** (0.427) |
| ... IPS | 7,017 | 3.774 [5.803] | -2.879*** (0.412) | -3.342*** (0.402) | -2.715*** (0.439) | -2.953*** (0.403) |
| ... SPS | 7,017 | 1.494 [3.399] | -0.413 (0.283) | -0.813*** (0.270) | -0.654** (0.281) | -0.577** (0.255) |
| ... Primary school | 7,017 | 2.053 [3.948] | 0.077 (0.202) | -0.181 (0.215) | 0.048 (0.218) | 0.004 (0.184) |

same as Table 6 for ever enrolled and total months enrolled by endline.

Table 8: Ever enrolled at endline by type of school and subgroups

| | Any preschool (1) | CPS (2) | IPS (3) | SPS (4) | Primary (5) |
|-------------------|----------------------------------|-----------------------------------|----------------------|------------------------------------|-------------------|
| Age 2 | 0.141*** (0.042) | 0.472*** (0.021) | -0.284*** (0.041) | -0.040* (0.022) | -0.001 (0.009) |
| Age 3 | 0.061* (0.035) | 0.550*** (0.024) | -0.313*** (0.048) | -0.105** (0.047) | 0.010 (0.027) |
| Age 4 | -0.006 (0.017) | 0.548*** (0.021) | -0.339*** (0.046) | -0.091** (0.042) | -0.029 (0.037) |
| Stunted | 0.056 (0.036) | 0.510*** (0.024) | -0.315*** (0.043) | -0.072** (0.034) | -0.015 (0.026) |
| Not stunted | 0.079*** (0.028) | 0.528*** (0.020) | -0.311*** (0.041) | -0.078** (0.036) | 0.017 (0.026) |
| Wealth Q1 | 0.052 (0.053) | 0.455*** (0.028) | -0.326*** (0.065) | -0.060 (0.043) | 0.003 (0.036) |
| Wealth Q2 | 0.095** (0.038) | 0.520*** (0.024) | -0.297*** (0.047) | -0.055 (0.041) | -0.004 (0.031) |
| Wealth Q3 | 0.082*** (0.030) | 0.581*** (0.023) | -0.319*** (0.048) | -0.069* (0.041) | -0.046 (0.035) |
| Wealth Q4 | 0.041 (0.034) | 0.540*** (0.026) | -0.305*** (0.051) | -0.138*** (0.046) | 0.047 (0.037) |
| Household educ Q1 | 0.065 (0.043) | 0.506*** (0.024) | -0.323*** (0.054) | -0.057 (0.035) | -0.005 (0.034) |
| Household educ Q2 | 0.088** (0.042) | 0.538*** (0.024) | -0.317*** (0.053) | -0.064 (0.039) | -0.001 (0.034) |
| Household educ Q3 | 0.085** (0.034) | 0.535*** (0.023) | -0.288*** (0.045) | -0.077* (0.043) | -0.023 (0.033) |
| Household educ Q4 | 0.032 (0.035) | 0.509*** (0.025) | -0.313*** (0.046) | -0.110** (0.044) | 0.025 (0.039) |
| Child cog. Q1 | 0.082** (0.040) | 0.435*** (0.025) | -0.290*** (0.044) | -0.042 (0.026) | -0.000 (0.016) |
| Child cog. Q2 | 0.120** (0.047) | 0.542*** (0.023) | -0.306*** (0.052) | -0.070* (0.040) | -0.018 (0.028) |
| Child cog. Q3 | 0.045 (0.040) | 0.556*** (0.024) | -0.321*** (0.055) | -0.075* (0.045) | -0.025 (0.040) |
| Child cog. Q4 | 0.029 (0.021) | 0.568*** (0.022) | -0.339*** (0.045) | -0.124*** (0.045) | 0.045 (0.038) |
| Female | 0.076** (0.032) | 0.535*** (0.021) | -0.328*** (0.042) | -0.068* (0.035) | -0.012 (0.027) |
| Male | 0.059** (0.029) | 0.509*** (0.020) | -0.294*** (0.039) | -0.087** (0.034) | 0.007 (0.025) |

Same as Table 6 for ever enrolled and by sub-groups formed from baseline variables. In bold, we indicate the subgroup estimates that are significantly different from the first category (at 10% level) e.g. children aged 3 or 4 are more likely to have ever enrolled to CPS than children aged 2.

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 9: ITT effects

| | Early numeracy | Language | Executive functions | Fine motor | Cognitive dev. index | Socio emotional index |
|-------------------------------|-------------------|------------------|---------------------|---------------------|----------------------|-----------------------|
| Midline data - age 3–5 | | | | | | |
| T1 | 0.051* (0.029) | 0.043 (0.030) | 0.043 (0.027) | 0.055* (0.030) | 0.054** (0.025) | 0.048 (0.038) |
| T2 | 0.012 (0.035) | 0.022 (0.038) | 0.005 (0.030) | 0.003 (0.035) | 0.010 (0.030) | 0.071* (0.043) |
| T3 | 0.062* (0.033) | 0.038 (0.033) | 0.041 (0.030) | 0.044 (0.030) | 0.052* (0.027) | 0.106** (0.043) |
| Any T | 0.044* (0.026) | 0.036 (0.028) | 0.033 (0.024) | 0.039 (0.027) | 0.042* (0.023) | 0.069** (0.035) |
| | 6917 | 6917 | 6917 | 6917 | 6917 | 6917 |
| Endline data - age 4–6 | | | | | | |
| T1 | -0.001 (0.028) | 0.045 (0.031) | -0.004 (0.026) | -0.017 (0.033) | 0.003 (0.026) | -0.007 (0.038) |
| T2 | -0.003 (0.032) | 0.007 (0.039) | -0.044 (0.030) | -0.076** (0.035) | -0.042 (0.030) | 0.015 (0.043) |
| T3 | 0.027 (0.034) | 0.031 (0.038) | 0.028 (0.032) | -0.042 (0.035) | 0.003 (0.031) | -0.007 (0.043) |
| Any T | 0.006 (0.025) | 0.032 (0.028) | -0.005 (0.022) | -0.038 (0.029) | -0.008 (0.022) | -0.001 (0.036) |
| Midline vs Endline | | | | | | |
| Any T - pvalue | 0.121 | 0.757 | 0.071 | 0.003 | 0.001 | 0.063 |
| Observations | 6966 | 6966 | 6966 | 6966 | 6966 | 6966 |

All regressions control for baseline value of dependent variable, child age, child age squared, and province fixed effects. In the panel *Midline vs Endline* we provide the p-value of the difference between midline and endline (any) treatment effect. Standard errors clustered on village level.

*, **, and *** indicate significance at the 10%, 5%, and % 1% levels, respectively.

Table 10: $LATE_{hc}$ Bounds - One Year Results

| | Obs. | Bounds | | Narrow Bounds | |
|------------------------------------|------|---------------------|--------------------|--------------------|--------------------|
| | | Lower | Upper | Lower | Upper |
| Any School Exposure (m) | 4006 | 2.708*** (0.822) | 9.2*** (1.605) | 2.184** (0.876) | 7.48*** (1.142) |
| Early numeracy | 3959 | 0.121* (0.071) | 0.407 (0.256) | 0.141* (0.077) | 0.194 (0.212) |
| Language | 3959 | 0.107 (0.076) | 0.36 (0.256) | 0.116 (0.079) | 0.265 (0.226) |
| Executive functions | 3959 | 0.146* (0.076) | 0.492* (0.26) | 0.154** (0.076) | 0.354 (0.238) |
| Fine motor | 3959 | 0.107* (0.065) | 0.359* (0.208) | 0.087 (0.064) | 0.335 (0.219) |
| <i>Cognitive development index</i> | 3959 | 0.138** (0.062) | 0.462** (0.212) | 0.138** (0.063) | 0.345* (0.198) |
| <i>Socio-emotional index</i> | 3959 | 0.118 (0.093) | 0.396 (0.333) | 0.134 (0.095) | 0.243 (0.307) |

Table 10 gives the bounds for the $LATE_{hc}$. The lower bound is the $LATE_{cps}$, the upper bound is the $LATE_{ps}$. $LATE_{cps}$ and $LATE_{ps}$ are estimated using province fixed effect, gender, age and baseline test scores as control variables (\mathbf{W}). Narrow bounds are estimated using \mathbf{W} and \mathbf{B} as control variables and instrument the endogenous variable (CPS enrollment or any preschool enrollment) by \mathbf{Z} , \mathbf{B} and $\mathbf{B}^*\mathbf{Z}$. \mathbf{B} is a dummy variable taking 1 when the predicted ASP enrollment is above the median. The predicted ASP is calculating using a set of variables selected via LASSO. The prediction is conducted on the sole control group. Standard errors are robust to heteroskedasticity and are clustered at the village level.

* 10%, ** 5%, *** 1% significance level

Table 11: SubLATEs - One year Results

| | Province FE | | | Prov. FE & vil- lage char. | | Prov. FE, vil- lage & hh char. | | Lasso (plug-in) | |
|--------------------------------|---------------------|---------------------------|--------------------|-------------------------------|-------------------|-----------------------------------|------------------|------------------------|-------------------|
| | $LATE_{cps}$ | $LATE_{hc}$ | $LATE_{ac}$ | $LATE_{hc}$ | $LATE_{ac}$ | $LATE_{hc}$ | $LATE_{ac}$ | $LATE_{hc}$ | $LATE_{ac}$ |
| Months of exposure | 2.708*** (0.822) | 8.88*** (0.992) | 0.333 (0.556) | 8.833*** (0.799) | 0.228 (0.528) | 8.771*** (0.741) | 0.225 (0.526) | 5.249*** (1) | -0.574 (0.529) |
| Overid. test p-value | | 0.496 | | 0.748 | | 0.891 | | . | |
| Early numeracy | 0.121* (0.071) | 0.015 (0.128) | 0.137 (0.094) | 0.107 (0.115) | 0.071 (0.092) | 0.199* (0.111) | 0.05 (0.089) | -0.114 (0.205) | 0.163* (0.099) |
| Overid. test p-value | | 0.82 | | 0.188 | | 0.092 | | . | |
| Language | 0.107 (0.076) | 0.021 (0.152) | 0.085 (0.098) | 0.141 (0.12) | 0.054 (0.095) | 0.197* (0.112) | 0.048 (0.089) | 0.233 (0.21) | 0.07 (0.093) |
| Overid. test p-value | | 0.613 | | 0.828 | | 0.743 | | . | |
| Fine motor | 0.146* (0.076) | -0.043 (0.181) | 0.197** (0.096) | 0.06 (0.163) | 0.166* (0.087) | 0.063 (0.153) | 0.16* (0.084) | 0.182 (0.214) | 0.116 (0.09) |
| Overid. test p-value | | 0.738 | | 0.854 | | 0.895 | | . | |
| Executive functions | 0.107* (0.065) | 0.045 (0.125) | 0.096 (0.081) | 0.12 (0.123) | 0.044 (0.074) | 0.225** (0.114) | 0.006 (0.073) | 0.335 (0.216) | 0.021 (0.072) |
| Overid. test p-value | | 0.2 | | 0.249 | | 0.044 | | . | |
| <i>Cognitive index</i> | 0.138** (0.062) | 0.012 (0.137) | 0.149* (0.08) | 0.119 (0.122) | 0.097 (0.075) | 0.192* (0.111) | 0.075 (0.071) | 0.219 (0.179) | 0.094 (0.069) |
| Overid. test p-value | | 0.722 | | 0.604 | | 0.23 | | . | |
| <i>Socio-emotional index</i> | 0.115 (0.094) | -0.011 (0.198) | 0.173 (0.142) | -0.002 (0.176) | 0.151 (0.127) | 0.135 (0.163) | 0.11 (0.129) | -0.046 (0.279) | 0.173* (0.097) |
| Overid. test p-value | | 0.386 | | 0.578 | | 0.316 | | . | |
| Weak instrument test (p-value) | | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 11 gives the $LATE_{cps}$ and the result of the conditional LATE approach for different set of \mathbf{X} instruments. The last columns uses as X variable the predicted ASP enrollment at midline calculated using the baseline variable selected by LASSO (plug-in). For each conditional LATE specification, we provide the p-value of the Sanderson-Windmeijer test of weak instruments and the p-value of the Sargan-Hansen over-identification test. The Estimates correct for heteroskedasticity and are clustered at the village level.

* 10%, ** 5%, *** 1% significance level

Table 12: Version 2: ITT effects by subgroups

| Baseline char. | Midline (age 3–5) | | Endline (age 4–6) | |
|-------------------|-----------------------------------|--------------------------|---------------------------------|--------------------------|
| | Cognitive development index | Socio emotional problems | Cognitive development index | Socio emotional problems |
| Age 2 | 0.023 (0.024) | 0.043 (0.052) | -0.028 (0.023) | 0.060 (0.060) |
| Age 3 | 0.057* (0.030) | 0.045 (0.047) | 0.021 (0.028) | -0.038 (0.054) |
| Age 4 | 0.033 (0.037) | 0.118** (0.051) | -0.019 (0.039) | -0.008 (0.057) |
| Stunted | -0.001 (0.025) | 0.047 (0.050) | -0.019 (0.025) | -0.015 (0.054) |
| Not stunted | 0.067*** (0.025) | 0.078* (0.041) | 0.003 (0.026) | 0.001 (0.041) |
| Wealth Q1 | 0.029 (0.034) | 0.111 (0.080) | -0.040 (0.030) | 0.020 (0.083) |
| Wealth Q2 | 0.005 (0.037) | 0.108** (0.050) | -0.015 (0.035) | 0.030 (0.056) |
| Wealth Q3 | 0.037 (0.034) | -0.004 (0.050) | -0.031 (0.036) | -0.072 (0.063) |
| Wealth Q4 | 0.054 (0.038) | 0.061 (0.068) | 0.049 (0.039) | -0.004 (0.066) |
| Household educ Q1 | 0.032 (0.038) | 0.098 (0.062) | -0.024 (0.033) | 0.062 (0.065) |
| Household educ Q2 | -0.034 (0.037) | -0.026 (0.062) | -0.049 (0.038) | -0.042 (0.067) |
| Household educ Q3 | 0.074** (0.032) | 0.105* (0.056) | -0.036 (0.032) | 0.009 (0.057) |
| Household educ Q4 | 0.084** (0.036) | 0.124* (0.065) | 0.075* (0.041) | -0.046 (0.065) |
| Child cog. Q1 | 0.004 (0.016) | 0.031 (0.064) | -0.044 (0.032) | -0.039 (0.072) |
| Child cog. Q2 | 0.012 (0.019) | 0.045 (0.057) | -0.041 (0.034) | -0.036 (0.057) |
| Child cog. Q3 | -0.010 (0.021) | 0.131** (0.059) | -0.016 (0.039) | 0.037 (0.063) |
| Child cog. Q4 | 0.007 (0.036) | 0.061 (0.049) | 0.052 (0.043) | 0.045 (0.058) |
| Female | 0.042 (0.026) | 0.052 (0.041) | -0.014 (0.029) | -0.009 (0.045) |
| Male | 0.037 (0.024) | 0.089** (0.043) | -0.001 (0.021) | 0.010 (0.045) |

All regressions control for individual baseline test scores, child age, province fixed effects and gender Standard errors are robust and clustered at the village level.

* 10%, ** 5%, *** 1 % significance level.

Table 13: Demand-side interventions take-up

| | Obs. | C | T1-C | T2-C | T3-C |
|--------------------------|------|-----------------|--------------------|--------------------|--------------------|
| Midline: | | | | | |
| Ever received home visit | 6552 | 0.688 (0.02) | 0.012 (0.02) | -0.018 (0.03) | 0.027 (0.03) |
| Ever received leaflet | 6552 | 0.053 (0.01) | 0.020* (0.01) | 0.056*** (0.01) | 0.079*** (0.01) |
| HBP Participation | 6552 | 0.176 (0.02) | 0.065*** (0.02) | 0.045 (0.03) | 0.190*** (0.03) |
| ... more than once | 6552 | 0.118 (0.02) | 0.040** (0.02) | 0.027 (0.02) | 0.133*** (0.03) |
| Endline: | | | | | |
| Ever received home visit | 6575 | 0.742 (0.02) | 0.001 (0.03) | 0.038 (0.03) | 0.03 (0.03) |
| Ever received leaflet | 6575 | 0.075 (0.01) | 0.008 (0.01) | 0.105*** (0.02) | 0.083*** (0.02) |
| HBP Participation | 6575 | 0.178 (0.02) | 0.056** (0.02) | 0.03 (0.02) | 0.100*** (0.03) |
| ... more than once | 6586 | 0.096 (0.01) | 0.047*** (0.02) | 0.035* (0.02) | 0.075*** (0.02) |

C is the control group mean and constant in a regression of the outcome variable on a set of dummy variables for each treatment group. No control variables included in regression model. Estimates correct for heteroskedasticity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix A: Program's content

Figure A1: Example of standardized CPS building and classroom.



Figure A2: English version of leaflet used in door-to-door intervention at baseline.

Community Preschool

Your village has been selected to benefit from an improved community preschool supported by the Cambodian Ministry of Education. To make sure that children have enough space to learn and thrive, the preschool will have its own building and equipment. A trained teacher will prepare children of age 3-5 for primary school for 2 hours per day. **If your child is of age 3-5, it is at the right age to benefit from the preschool. Preschool is free for all children.**

It is important for your child that it constantly learns new things. Preschool education can help children to become more intelligent and well-behaved. The community preschool is a place where children can learn how to interact with each other and learn about honesty, respect, sharing and perseverance. They will also learn about numbers, letters and words. Visiting a community preschool can help your child to stay in school longer and to do well in his/her future.

School education is very important for children. Data from the Cambodian Socioeconomic Survey 2009 has shown that children who stay in school longer are likely to earn more in the future.

Average Monthly Income by Education

| Education Level | Average Monthly Income (KHR) |
|----------------------------------|------------------------------|
| Primary school not completed | 260,000 |
| Primary school completed | 330,000 |
| Lower secondary school completed | 410,000 |
| Upper secondary school completed | 680,000 |

Figure A3: English version of leaflet used in door-to-door intervention at midline.

Community Preschool



A stimulating environment is crucial for optimal development of your child. Preschool education can contribute to a better future for your child by providing new learning experiences every day.

At preschool, a trained teacher will help your child to learn important values such as respect, sharing and perseverance. Children will also be prepared for primary school by learning about numbers, letters and words.

If you have a child age 3-5, you can enroll your child at preschool!

Appendix B: Additional Analysis

Figure B1: Enrollment before and After School Construction - C and T1 groups

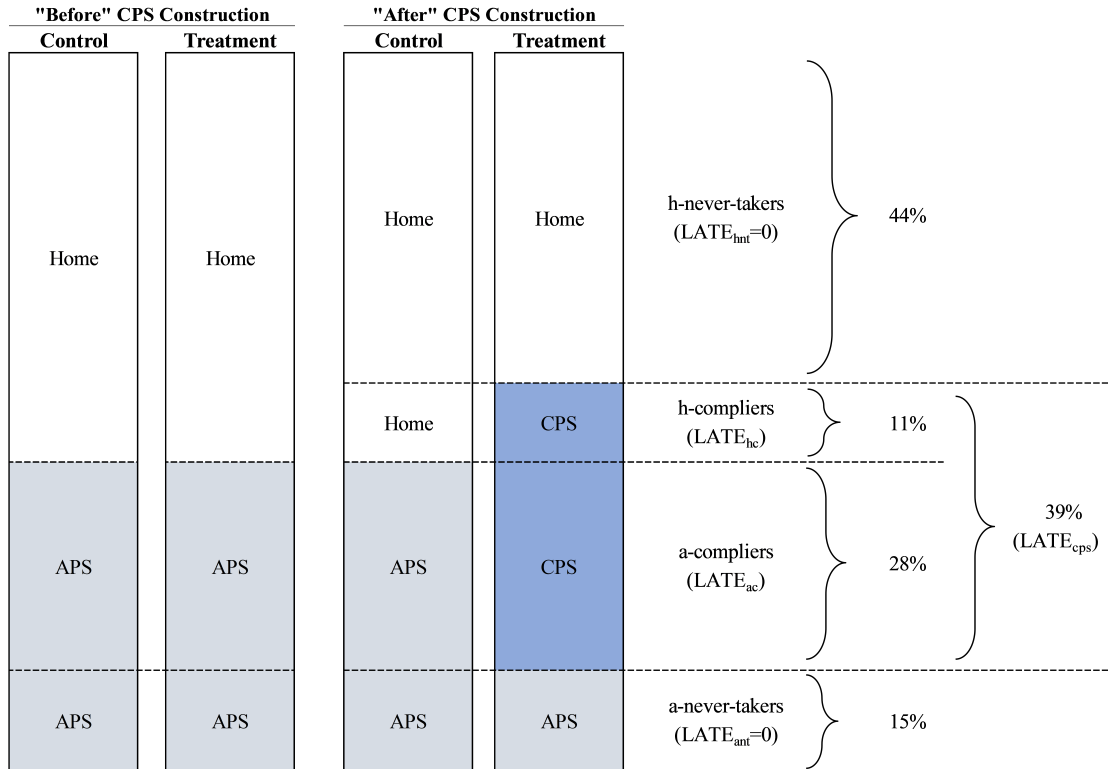


Figure B1 shows care arrangements ($D \in \{c, a, h\}$) of children in treatment and control groups. The left panel shows the counterfactual scenario in the absence of the program. The right panel shows the observed scenario at midline under implementation of the program. Randomization implies that the control group at midline is equivalent to the treatment group at midline in the absence of the program.

Table B1: Attrition of eligible children at midline and endline

| | Midline | | | Endline | | |
|----------------------------------|---------|--------|--------|---------|----------|--------|
| | (1) | (2) | (3) | (1) | (2) | (3) |
| T1 | 0.019 | -0.062 | 0.021* | -0.003 | 0.067 | -0.001 |
| T2 | 0.008 | 0.058 | 0.009 | 0.008 | 0.091 | 0.009 |
| T3 | -0.008 | 0.001 | -0.006 | -0.002 | 0.050 | 0.000 |
| Cognitive development index | | 0.006 | | | -0.010 | |
| Age | | -0.015 | | | 0.008 | |
| Height-for-age z-score | | -0.005 | | | -0.009 | |
| Multidim. poor | | 0.002 | | | 0.016 | |
| Caregiver education | | 0.001 | | | -0.001 | |
| T1 * Cognitive index | | -0.015 | | | 0.006 | |
| T1 * Age | | 0.014 | | | -0.020 | |
| T1 * Height-for-age z-score | | -0.007 | | | 0.006 | |
| T1 * Multidim. poor | | 0.028 | | | 0.001 | |
| T1 * Caregiver education | | 0.001 | | | 0.001 | |
| T2 * Cognitive index | | -0.002 | | | -0.006 | |
| T2 * Age | | -0.007 | | | -0.007 | |
| T2 * Height-for-age z-score | | 0.009 | | | 0.041*** | |
| T2 * Multidim. poor | | -0.011 | | | 0.019 | |
| T2 * Caregiver education | | -0.002 | | | -0.001 | |
| T3 * Cognitive index | | 0.001 | | | 0.008 | |
| T3 * Age | | 0.001 | | | -0.015 | |
| T3 * Height-for-age z-score | | 0.004 | | | 0.005 | |
| T3 * Multidim. poor | | -0.004 | | | 0.003 | |
| T3 * Caregiver education | | -0.001 | | | 0.001 | |
| Control group mean attrition | 0.102 | | | 0.0956 | | |
| Joint F-test (p-values): | | | | | | |
| Baseline control, no interaction | | 0.847 | | | 0.190 | |
| T1 with baseline interactions | | 0.525 | | | 0.423 | |
| T2 with baseline interaction | | 0.676 | | | 0.0434 | |
| T3 with baseline interaction | | 0.986 | | | 0.912 | |
| Additional sampling at midline | No | No | Yes | No | No | Yes |
| Observations | 7632 | 7632 | 7693 | 7632 | 7632 | 7693 |

Table shows OLS regressions with midline and endline attrition as dependent variable using all children eligible for testing at baseline (columns 1 and 2) plus children that were added to the sample at midline (column 3). All regressions also control for province fixed effects. Robust standard errors are clustered at village level. Missing baseline control variables are replaced by the control group mean. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B2: Teacher characteristics and preschool equipment

| | CPS (1) | IPS (2) | SPS (3) | IPS-CPS (4) | SPS-CPS (5) | N (6) |
|--|--------------|--------------|-------------|------------------|----------------|------------|
| Teacher characteristics | -0.37 | -0.22 | 1.41 | 0.148 | 1.781*** | 328 |
| Female | 0.94 | 0.97 | 0.84 | 0.03 | -0.10* | 329 |
| Age | 40.46 | 40.92 | 34.05 | 0.45 | -6.42*** | 329 |
| Years since first teaching experience | 6.04 | 6.46 | 5.97 | 0.41 | -0.08 | 329 |
| Completed primary school, 6 y | 0.81 | 0.78 | 0.98 | -0.03 | 0.18*** | 329 |
| Completed lower secondary school, 9 y | 0.43 | 0.39 | 0.86 | -0.04 | 0.42*** | 329 |
| Completed upper secondary school, 12 y | 0.13 | 0.14 | 0.68 | 0.01 | 0.55*** | 329 |
| No teacher training | 0.12 | 0.15 | 0.29 | 0.04 | 0.17*** | 329 |
| 1-4 weeks of teacher training | 0.13 | 0.32 | 0.27 | 0.19*** | 0.14** | 329 |
| 5-8 weeks of teacher training | 0.70 | 0.46 | 0.03 | -0.24*** | -0.67*** | 329 |
| > 8 weeks of training | 0.05 | 0.07 | 0.41 | 0.01 | 0.36*** | 329 |
| Had practical teacher training | 0.21 | 0.20 | 0.41 | -0.00 | 0.20*** | 329 |
| Trained as prim./sec. school teacher | 0.02 | 0.05 | 0.44 | 0.03 | 0.42*** | 329 |
| Nonverbal reasoning test (Raven's) | -0.10 | -0.04 | 0.41 | 0.05 | 0.51*** | 328 |
| Salary for teaching position (USD) | 60.74 | 67.32 | 250.22 | 6.58 | 189.48*** | 329 |
| Teacher fully paid, regularly | 0.75 | 0.85 | 0.95 | 0.10* | 0.20*** | 329 |
| Teacher fully paid, irregularly | 0.22 | 0.15 | 0.03 | -0.07 | -0.19*** | 329 |
| Equipment | 0.08 | -0.51 | 0.22 | -0.592*** | 0.138 | 326 |
| Table and chair for teacher | 0.99 | 0.34 | 0.87 | -0.65*** | -0.12*** | 329 |
| Storage for teacher | 0.97 | 0.19 | 0.52 | -0.78*** | -0.44*** | 329 |
| Tables and chairs for children | 0.95 | 0.32 | 0.67 | -0.63*** | -0.29*** | 329 |
| Tables & chairs appropriately sized | 0.78 | 0.29 | 0.34 | -0.48*** | -0.44*** | 327 |
| Board & markers | 0.95 | 0.83 | 0.97 | -0.12** | 0.02 | 329 |
| Electricity access | 0.07 | 0.25 | 0.24 | 0.19*** | 0.17*** | 329 |
| Field, playground or school yard | 0.62 | 0.61 | 0.83 | -0.01 | 0.21*** | 329 |
| Equipment for gross-motor in school | 0.38 | 0.32 | 0.30 | -0.05 | -0.08 | 329 |
| First aid kit | 0.31 | 0.12 | 0.33 | -0.20*** | 0.02 | 329 |
| Functional water source | 0.46 | 0.64 | 0.81 | 0.18** | 0.35*** | 329 |
| Functional drinking water source | 0.61 | 0.54 | 0.67 | -0.07 | 0.05 | 329 |
| Hand washing facility | 0.54 | 0.37 | 0.57 | -0.16** | 0.04 | 329 |
| Toilet facility | 0.27 | 0.42 | 0.90 | 0.15** | 0.63*** | 329 |
| Writing utensils | 0.94 | 0.80 | 0.89 | -0.14** | -0.05 | 329 |
| Writing utensils used by children | 0.61 | 0.63 | 0.79 | 0.01 | 0.18*** | 329 |
| Art materials | 0.91 | 0.71 | 0.81 | -0.20*** | -0.10* | 329 |
| Art materials used by children | 0.56 | 0.53 | 0.70 | -0.03 | 0.14** | 329 |
| Fantasy play materials | 0.65 | 0.29 | 0.25 | -0.36*** | -0.40*** | 329 |
| Children use fantasy play materials | 0.40 | 0.19 | 0.08 | -0.21*** | -0.32*** | 329 |
| Educational toys/math materials | 0.77 | 0.44 | 0.67 | -0.33*** | -0.10 | 329 |
| Children use Educational materials | 0.43 | 0.25 | 0.46 | -0.18*** | 0.03 | 329 |
| Number of schools | 207 | 59 | 63 | | | |

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Differences are based on regressions of dependent variables on binary preschool type variables using robust standard errors. Summary scores are prepared using the first principal component of the individual variables and are standardized to a mean of zero and a standard deviation of one. Observations with missing values are dropped from the summary scores. Variables with a *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B3: Classroom setting and teaching practices

| | CPS (1) | IPS (2) | SPS (3) | IPS-CPS (4) | SPS-CPS (5) | N (6) |
|--|--------------|--------------|--------------|------------------|------------------|------------|
| Classroom setting | 0.35 | -1.07 | -0.18 | -1.419*** | -0.539*** | 327 |
| Length of class (min) | 113.78 | 137.53 | 172.49 | 23.74 | 58.71*** | 328 |
| Total length of breaks (min) | 43.22 | 21.07 | 44.22 | -22.15* | 1.00 | 329 |
| Number of children enrolled in this class | 25.09 | 22.56 | 27.57 | -2.53** | 2.48** | 329 |
| Children present | 17.66 | 14.98 | 21.21 | -2.68** | 3.54*** | 329 |
| Num. of teachers in classroom | 1.01 | 0.98 | 0.98 | -0.03 | -0.03 | 329 |
| Num. of assistants in classroom | 0.03 | 0.03 | 0.00 | 0.00 | -0.03** | 329 |
| Num. of other adults in classroom | 1.03 | 0.73 | 0.31 | -0.30 | -0.72*** | 327 |
| Teacher follows curriculum to teach class | 0.68 | 0.46 | 0.49 | -0.22*** | -0.18** | 329 |
| Teacher documents children's development | 0.38 | 0.19 | 0.37 | -0.19*** | -0.01 | 329 |
| Teacher documents attendance | 0.87 | 0.68 | 0.86 | -0.20*** | -0.02 | 329 |
| Curriculum content and pedagogy | -0.05 | -0.23 | 0.38 | -0.179 | 0.428*** | 329 |
| Activities supporting maths [×] | 0.82 | 0.80 | 0.71 | -0.02 | -0.10 | 329 |
| Quality of maths activities [1-4] | 2.68 | 2.47 | 2.89 | -0.21 | 0.21 | 259 |
| Activities supporting literacy [×] | 0.73 | 0.76 | 0.89 | 0.03 | 0.16*** | 329 |
| Quality of literacy activities [1-4] | 2.66 | 2.47 | 3.39 | -0.19 | 0.74*** | 249 |
| Activities supporting expressive language [×] | 0.88 | 0.80 | 0.78 | -0.09 | -0.11* | 329 |
| Quality of expressive language activities [1-3] | 2.37 | 2.28 | 2.67 | -0.09 | 0.31** | 279 |
| Activity: reading of storybook [×] | 0.54 | 0.47 | 0.46 | -0.07 | -0.08 | 329 |
| Quality of storybook activities [1-6] | 3.54 | 2.93 | 3.86 | -0.62* | 0.31 | 168 |
| Activities supporting general knowledge [×] | 0.85 | 0.90 | 0.87 | 0.05 | 0.03 | 329 |
| Teaching quality, general knowledge act. [1-6] | 3.78 | 3.89 | 4.00 | 0.10 | 0.22 | 283 |
| Activities supporting fine motor skills [×] | 0.39 | 0.41 | 0.57 | 0.02 | 0.18** | 329 |
| Teaching quality in fine motor skills act. [1-3] | 2.04 | 2.09 | 2.31 | 0.05 | 0.27** | 140 |
| Activities supporting gross motor skills [×] | 0.67 | 0.58 | 0.75 | -0.09 | 0.08 | 329 |
| Quality of gross motor skills activities [1-3] | 1.63 | 1.56 | 1.85 | -0.07 | 0.22 | 215 |
| Time of gross motor skills activities (min) [×] | 7.87 | 8.75 | 10.42 | 0.88 | 2.55** | 129 |
| Quality of the teacher's use of theme [0-4] [×] | 2.06 | 2.13 | 2.17 | 0.06 | 0.11 | 280 |
| Teacher-child interactions | -0.06 | -0.10 | 0.29 | -0.044 | 0.347** | 321 |
| The teacher enjoyed teaching [1-3] | 2.58 | 2.49 | 2.60 | -0.09 | 0.02 | 329 |
| The teacher showed negative attitudes [0-2.5] | 0.06 | 0.07 | 0.06 | 0.00 | 0.00 | 329 |
| Quality of the disciplinary strategies [0-4] | 2.95 | 3.02 | 3.13 | 0.07 | 0.18 | 329 |
| > 4x negative interactions | 0.44 | 0.53 | 0.51 | 0.09 | 0.07 | 324 |
| > 8x encouragements | 0.43 | 0.44 | 0.60 | 0.01 | 0.17** | 326 |
| Children wait > 5 min without any activity | 0.15 | 0.10 | 0.10 | -0.05 | -0.06 | 329 |
| The teacher correct work & give feedbacks [0-3] | 1.86 | 2.00 | 2.32 | 0.14 | 0.46*** | 329 |
| Children ever left without supervision | 0.14 | 0.20 | 0.24 | 0.06 | 0.09 | 329 |
| Quality of the children engagement [0-4] | 2.57 | 2.76 | 3.25 | 0.20 | 0.69*** | 329 |
| Teacher's awareness of children's needs [0-3] | 1.41 | 1.36 | 1.44 | -0.05 | 0.04 | 329 |
| Teacher's respect to gender equality [0-4] | 3.26 | 3.19 | 3.35 | -0.08 | 0.09 | 328 |
| Class is interrupted at least once | 0.58 | 0.58 | 0.62 | -0.00 | 0.04 | 329 |
| Presence of disturbing noise [1-3] | 1.36 | 1.47 | 1.27 | 0.11 | -0.09 | 329 |
| Number of schools | 207 | 59 | 63 | | | |

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Differences are based on regressions of dependent variables on binary preschool type variables using robust standard errors. Summary scores are prepared using the first principal component of the individual variables and are standardized to a mean of zero and a standard deviation of one. Observations with missing values are dropped from the summary scores. Variables marked with [×] were dropped from the principal component analysis to avoid missing values in summary scores. Activities which were not taught were assigned a quality score of 0. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B4: Perceived preschool quality and financial contributions of parents

| | CPS (1) | IPS (2) | SPS (3) | IPS-CPS (4) | SPS-CPS (5) | N (6) |
|--|------------|------------|------------|----------------------|----------------------|----------|
| Midline | | | | | | |
| Regular school days without class, last 30 d | 2.48 | 1.51 | 0.92 | -0.969*** (0.321) | -1.556*** (0.305) | 3378 |
| Days missed due to personal reasons (last 30 days) | 4.32 | 3.51 | 3.42 | -0.813** (0.332) | -0.902*** (0.332) | 3378 |
| Perceived kindness of teacher (1-10) | 8.74 | 8.25 | 8.57 | -0.482*** (0.130) | -0.164* (0.090) | 3379 |
| Perceived professional knowledge of teacher (1-10) | 8.62 | 8.09 | 8.56 | -0.523*** (0.127) | -0.060 (0.093) | 3379 |
| Perceived reliability of teacher (1-10) | 8.38 | 8.02 | 8.43 | -0.364** (0.141) | 0.049 (0.122) | 3379 |
| Contribution to teacher salary, USD | 0.14 | 0.42 | 0.16 | 0.281** (0.139) | 0.022 (0.069) | 3360 |
| Contribution to school material, USD | 35.15 | 31.70 | 49.22 | -3.440 (2.641) | 14.080*** (2.748) | 3354 |
| Contribution to construction, USD | 0.67 | 0.67 | 0.64 | -0.000 (0.215) | -0.029 (0.135) | 3345 |
| Endline | | | | | | |
| Regular school days without class, last 30 d) | 2.13 | 1.78 | 1.31 | -0.354 (0.334) | -0.818*** (0.216) | 3193 |
| Days missed due to personal reasons (last 30 days) | 3.47 | 2.82 | 3.15 | -0.649** (0.280) | -0.313 (0.309) | 3050 |
| Perceived kindness of teacher (1-10) | 8.87 | 8.71 | 8.69 | -0.161* (0.093) | -0.183** (0.074) | 3193 |
| Perceived professional knowledge of teacher (1-10) | 8.72 | 8.57 | 8.80 | -0.153* (0.083) | 0.073 (0.070) | 3193 |
| Perceived reliability of teacher (1-10) | 8.65 | 8.64 | 8.84 | -0.014 (0.116) | 0.190** (0.078) | 3193 |
| Contribution to teacher salary, USD | 0.18 | 0.96 | 0.12 | 0.776*** (0.251) | -0.058 (0.066) | 3158 |
| Contribution to school material, USD | 46.74 | 53.32 | 68.29 | 6.578** (3.155) | 21.549*** (2.468) | 3165 |
| Contribution to construction, USD | 0.74 | 0.77 | 0.87 | 0.027 (0.185) | 0.129 (0.161) | 3172 |

Contributions are trimmed at the 99th percentile to control for outliers. Columns 1—3 show averages by type of preschool. Columns 4—5 show differences between types of preschools. Robust standard errors clustered at village level in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B5: Months enrolled at endline by type of school and subgroups

| | Any preschool (1) | CPS (2) | IPS (3) | SPS (4) | Primary (5) |
|--------------------|-----------------------------------|-----------------------------------|------------------------------------|------------------------------------|--------------------------------|
| Age 2 | 1.028*** (0.391) | 3.644*** (0.187) | -2.426*** (0.363) | -0.190 (0.146) | -0.076 (0.080) |
| Age 3 | 0.768 (0.531) | 4.699*** (0.227) | -3.023*** (0.534) | -0.907** (0.380) | 0.131 (0.199) |
| Age 4 | 1.025* (0.539) | 5.166*** (0.230) | -3.469*** (0.538) | -0.671* (0.359) | -0.125 (0.359) |
| Stunted | 0.818 (0.500) | 4.313*** (0.211) | -3.086*** (0.488) | -0.408 (0.250) | -0.253 (0.227) |
| Not stunted | 0.965** (0.455) | 4.631*** (0.201) | -2.988*** (0.427) | -0.679** (0.303) | 0.206 (0.213) |
| Wealth Q1 | 0.252 (0.680) | 3.727*** (0.244) | -3.208*** (0.725) | -0.267 (0.302) | -0.005 (0.275) |
| Wealth Q2 | 1.438*** (0.524) | 4.502*** (0.227) | -2.753*** (0.471) | -0.311 (0.270) | 0.015 (0.279) |
| Wealth Q3 | 1.484** (0.600) | 5.063*** (0.249) | -3.116*** (0.564) | -0.463 (0.343) | -0.422 (0.309) |
| Wealth Q4 | 0.566 (0.551) | 4.752*** (0.256) | -2.789*** (0.543) | -1.397*** (0.418) | 0.554 (0.343) |
| Household educ. Q1 | 0.572 (0.646) | 4.300*** (0.236) | -3.392*** (0.630) | -0.336 (0.252) | 0.036 (0.273) |
| Household educ. Q2 | 1.038 (0.665) | 4.521*** (0.244) | -3.107*** (0.662) | -0.375 (0.262) | -0.133 (0.313) |
| Household educ. Q3 | 1.545*** (0.496) | 4.575*** (0.223) | -2.417*** (0.427) | -0.613* (0.356) | -0.042 (0.274) |
| Household educ. Q4 | 0.550 (0.564) | 4.593*** (0.246) | -3.035*** (0.500) | -1.007** (0.410) | 0.209 (0.349) |
| Child cog. Q1 | 0.674 (0.444) | 3.487*** (0.202) | -2.696*** (0.442) | -0.118 (0.170) | -0.085 (0.152) |
| Child cog. Q2 | 1.095* (0.604) | 4.393*** (0.230) | -2.800*** (0.582) | -0.498* (0.282) | -0.202 (0.228) |
| Child cog. Q3 | 1.329** (0.587) | 4.838*** (0.236) | -2.886*** (0.593) | -0.623* (0.370) | 0.033 (0.295) |
| Child cog. Q4 | 0.714 (0.626) | 5.404*** (0.260) | -3.567*** (0.520) | -1.123*** (0.421) | 0.373 (0.392) |
| Female | 1.062** (0.494) | 4.756*** (0.214) | -3.159*** (0.466) | -0.535* (0.281) | 0.022 (0.232) |
| Male | 0.859** (0.425) | 4.216*** (0.184) | -2.738*** (0.419) | -0.619** (0.269) | 0.002 (0.199) |

Same as Table 6 for months enrolled and by sub-groups formed from baseline variables. In bold, we indicate the subgroup estimates that are significantly different from the first category (at 10% level) e.g. children aged 3 or 4 are more likely to have ever enrolled to CPS than children aged 2.

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B6: Impact of the program on parenting domains

| | obs | T1 | T2 | T3 | Any T |
|--------------------------|-------|-------------------|-------------------|--------------------|-------------------|
| Midline (age 3–5) | | | | | |
| Negative Parenting | 6,993 | 0.014 (0.037) | 0.026 (0.045) | 0.033 (0.044) | 0.022 (0.033) |
| Socioemotional Parenting | 6,993 | 0.014 (0.045) | 0.021 (0.050) | 0.061 (0.049) | 0.028 (0.041) |
| Cognitive parenting | 6,993 | 0.065 (0.041) | 0.032 (0.043) | 0.100** (0.044) | 0.066* (0.036) |
| Endline (age 4–6) | | | | | |
| Negative Parenting | 6,963 | -0.006 (0.042) | 0.038 (0.043) | -0.020 (0.048) | 0.002 (0.039) |
| Socioemotional Parenting | 6,963 | 0.071* (0.042) | 0.000 (0.050) | 0.065 (0.049) | 0.052 (0.040) |
| Cognitive parenting | 6,963 | -0.015 (0.040) | -0.011 (0.048) | 0.017 (0.049) | -0.006 (0.037) |

All regressions control for baseline value of dependent variable, child age, child age squared, and province fixed effects. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B7: Impact of the program on parenting domains by subgroups

| Baseline characteristics | Midline (age 3–5) | | | Endline (age 4–6) | | |
|--------------------------|---------------------|---------------------------|---------------------------------|---------------------|---------------------------------|---------------------------------|
| | Cognitive parenting | Socio-emotional parenting | Negative parenting | Cognitive parenting | Socio-emotional parenting | Negative parenting |
| Age 2 | 0.040 (0.055) | 0.001 (0.068) | 0.027 (0.057) | 0.004 (0.052) | 0.039 (0.055) | -0.026 (0.060) |
| Age 3 | 0.063 (0.054) | 0.036 (0.054) | 0.025 (0.053) | 0.021 (0.048) | 0.044 (0.053) | 0.041 (0.049) |
| Age 4 | 0.093 (0.059) | 0.030 (0.067) | 0.025 (0.054) | -0.044 (0.067) | 0.051 (0.065) | -0.013 (0.058) |
| Stunted | 0.059 (0.057) | 0.037 (0.065) | 0.004 (0.052) | 0.015 (0.051) | 0.063 (0.074) | 0.047 (0.059) |
| Not stunted | 0.062 (0.044) | 0.024 (0.042) | 0.037 (0.041) | -0.019 (0.044) | 0.045 (0.039) | -0.009 (0.043) |
| Wealth Q1 | 0.046 (0.061) | 0.008 (0.081) | 0.044 (0.068) | -0.044 (0.060) | 0.060 (0.076) | -0.041 (0.073) |
| Wealth Q2 | 0.086 (0.055) | 0.060 (0.061) | -0.039 (0.059) | 0.048 (0.061) | 0.048 (0.065) | 0.045 (0.068) |
| Wealth Q3 | -0.055 (0.079) | -0.002 (0.073) | 0.051 (0.068) | -0.062 (0.067) | -0.033 (0.068) | -0.008 (0.067) |
| Wealth Q4 | 0.132* (0.080) | 0.026 (0.065) | 0.052 (0.074) | -0.006 (0.075) | 0.082 (0.065) | 0.045 (0.071) |
| Household educ Q1 | 0.070 (0.059) | 0.051 (0.068) | 0.056 (0.066) | 0.058 (0.054) | 0.264*** (0.070) | 0.053 (0.072) |
| Household educ Q2 | -0.058 (0.065) | 0.076 (0.075) | 0.091 (0.066) | -0.093 (0.078) | 0.019 (0.067) | 0.033 (0.069) |
| Household educ Q3 | 0.103 (0.063) | -0.011 (0.069) | -0.006 (0.059) | 0.099* (0.059) | -0.038 (0.057) | 0.007 (0.073) |
| Household educ Q4 | 0.149* (0.078) | -0.030 (0.061) | -0.054 (0.068) | -0.065 (0.081) | -0.071 (0.056) | -0.057 (0.067) |
| Child cog. Q1 | 0.046 (0.054) | 0.066 (0.083) | -0.099 (0.067) | -0.007 (0.061) | 0.004 (0.057) | -0.103 (0.067) |
| Child cog. Q2 | -0.007 (0.060) | -0.061 (0.069) | 0.020 (0.057) | -0.032 (0.068) | 0.034 (0.070) | 0.082 (0.067) |
| Child cog. Q3 | 0.039 (0.073) | -0.001 (0.074) | 0.046 (0.063) | 0.031 (0.061) | 0.051 (0.069) | 0.079 (0.063) |
| Child cog. Q4 | 0.135* (0.081) | 0.066 (0.064) | 0.091* (0.053) | -0.022 (0.074) | 0.053 (0.068) | -0.031 (0.061) |
| Female | 0.051 (0.048) | 0.045 (0.056) | 0.053 (0.043) | -0.040 (0.052) | 0.049 (0.052) | 0.077 (0.052) |
| Male | 0.072 (0.045) | -0.002 (0.047) | -0.009 (0.047) | 0.029 (0.044) | 0.046 (0.051) | -0.070 (0.047) |

Tables shows estimates from separate regressions of an outcome variables on a joint treatment group (T1-T3) dummy variable and control variables. Education background refers to the average years household members spent in school. All regressions include usual control variables. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. In bold, we indicate which estimate are significantly different (at 10%) from the first category. For instance, negative parenting is significantly worse for male than female children. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table B8: Perceived return to education

| | Optimal preschool age (1) | Optimal primary school age (2) | Return primary school (3) | Return secondary school (4) |
|---------------------------|------------------------------------|---|------------------------------------|--------------------------------------|
| Midline: | | | | |
| T1 | -0.139** (0.054) | -0.051* (0.027) | 0.005 (0.004) | 0.000 (0.003) |
| T2 | -0.139** (0.059) | -0.027 (0.028) | 0.007* (0.004) | 0.004 (0.004) |
| T3 | -0.219*** (0.060) | -0.058** (0.029) | 0.002 (0.004) | 0.001 (0.004) |
| <i>Control group mean</i> | <i>4.124</i> | <i>5.942</i> | <i>0.065</i> | <i>0.086</i> |
| T2 or T3 | -0.180*** (0.052) | -0.043* (0.025) | 0.005 (0.004) | 0.002 (0.003) |
| Observations | 6611 | 6611 | 6611 | 6611 |
| Endline: | | | | |
| T1 | -0.122** (0.047) | -0.051* (0.029) | 0.003 (0.004) | 0.000 (0.003) |
| T2 | -0.126** (0.056) | 0.012 (0.035) | 0.004 (0.004) | 0.002 (0.003) |
| T3 | -0.157*** (0.049) | -0.070** (0.032) | 0.000 (0.004) | 0.001 (0.003) |
| <i>Control group mean</i> | <i>4.394</i> | <i>6.080</i> | <i>0.068</i> | <i>0.085</i> |
| T2 or T3 | -0.142*** (0.046) | -0.030 (0.029) | 0.002 (0.003) | 0.002 (0.003) |
| Observations | 7014 | 7014 | 7014 | 7014 |

We use province dummy and age as control variables. Row “T2 or T3” shows parameter of regression model with a joint dummy variable for T2 and T3. Estimates correct for heteroskedasticity and within-village correlations. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Appendix C: Children test scores

This Annex summarizes the individual tests and scoring methods used in this paper. An in-depth discussion of the tests, scoring methods, cultural adaptations and pretesting procedures can be found in Berkes et al. (2019).

To ensure that children correctly understood the tests and that the test were reliable, the research team pretested every instrument at least three times before collecting data in the sample villages. The survey firm translated the questionnaires into Khmer and an independent third party back-translated them into English which led to further refinements in the instruments. The final child assessments included a total of 15 individual tests at baseline, 17 at midline and 20 at endline.

Before constructing the composite scores of child test domains, individual tests were first scored and standardized thus ensuring that all tests contributed equal variance to their composite score. Scoring was done by assigning 1 point for each correct response and summing up these points to create an individual score for each test. When a child was unable to complete the practice trial of a test, a score of zero was assigned for this test as long as the child participated in the other tests. Standardization of each test score was done with the control group mean of the same wave by subtracting the mean and dividing by the standard deviation of this wave. All standardized test scores of one domain (e.g. executive function) were then aggregated into a domain score (either cognitive development score or socio-emotional development score) using the approach described in Anderson (2008) and standardized again by subtracting its sample mean and dividing by the sample standard deviation of the domain score for better interpretability. After these steps, we obtained the following composite scores:

1. **Executive function:**

- 1.1. The construct inhibitory control is assessed with the head-knee task. The test has two stages. In the first stage, the child stands in front of the enumerator and is asked five times to touch his/her head or knees. In a second stage, the child is asked to do the opposite of what the enumerator says.
- 1.2. Working memory (short-term auditory memory) is assessed with a forward digit span test in which children have to repeat sequences of digits which increase in length.
- 1.3. The Dimensional Change Card Sort test is used as a measure of cognitive flexibility. We followed the procedures outlined in Zelazo (2006) using cards with two colours (blue and red) and two pictures (boat and rabbit). To reduce the burden on tested children, we followed the protocol with the exception that children needed to pass the pre-switch phase (at least 5 out of 6 correct) in order to participate in the post switch phase. The border

version of the test was only administered at endline. The demonstration phase of the test included one practice trial. As per protocol, this practice trial was not used to determine whether a child is eligible for the test as it could have performed well by chance.

- 1.4. We use a self-developed cancellation task to measure sustained attention. In this test, children see a printed matrix with different symbols and are asked to cross-out all symbols that match the given one (e.g. cross out all flowers). When completed, a larger matrix is given and a new symbol has to be crossed-out. The test continues until a child has completed 4 matrices, crossed out more wrong than correct images in a matrix, until the child loses attention, or states that it is done. The test was scored by using the difference between correctly and incorrectly crossed out images.

2. Language:

- 2.1. Receptive vocabulary skills are assessed with a test derived from the TVIP. In this test children are asked to match a word to one out of four pictures. The version used in the Cambodian context was culturally adapted during piloting and validation exercises prior to baseline data collection and with the support of key informants. The final instrument includes 82 pictures with a rule that the test stops after 6 out of the last 8 pictures were wrong. All other language development tests were taken from the MELQO.
- 2.2. Expressive language skills are assessed by asking children to name up to 10 things that can be eaten and up to 10 animals they know. The final score is the number of recalled items.
- 2.3. Receptive language is assessed with a listening comprehension test in which a short story (116 words) is read to the child. After reading the story, the child is asked five questions about the content of the story.
- 2.4. Knowledge of reading concepts is assessed by showing a children's story-book and asking how the book should be opened and where and in which direction one should start reading the story.
- 2.5. Reading skills are assessed with a letter name knowledge test in which children have to identify common letters of Khmer script.
- 2.6. *Endline only:* A name writing test was conducted to assess whether children were able to write their own name.
- 2.7. *Endline only:* An initial letter identification test was conducted in which children were asked to name the first alphabet letter of words that were read to the child

2.8. *Endline only*: Reading skills were assessed by asking the child to read out loud different printed words.

3. **Early numeracy:**

3.1. *Midline only*: The tests for early numeracy includes a self-developed test for measurement concepts, e.g. if the child understands concepts such as tallest/shortest, in which the child had to point to different printed objects.

3.2. In a test for verbal counting, children had to count up to 30.

3.3. Numbers and operations are also administered with a self-developed quantitative comparison test where children had to compare the number of printed objects on two sides of a page.

3.4. A number identification test analogous to the letter name knowledge test was used.

3.5. A self-developed shape recognition test was used to test if children are able to identify basic geometric shapes.

3.6. *Endline only*: Children were asked to read printed arithmetic problems and say the correct answer (e.g. $2+1$).

3.7. *Endline only*: A spacial vocabulary test was conducted in which the child was shown 4 pictures with a ball and a chair. The child was asked to point to the correct picture with the ball either on, under, in front or next to the chair.

4. **Fine-motor development:**

4.1. A drawing test, where children copy shapes, like circles or squares, was used to assess fine-motor skills.

4.2. A draw-a-person test

5. **Socio-emotional problems:** The recommended method was used to create a total difficulties score, i.e. summing up scores of the individual subcomponents without standarizing first. The subcomponents are:

5.1. Emotional symptoms

5.2. Conduct problems

5.3. Hyperactivity/inattention

5.4. Peer problems

6. **MDAT socio-emotional skills:** We use the socio-emotional MDAT test score for socio-emotional skills