Title
Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds, Conditional LATE and Machine Learning

Authors
Berkes, Jan
Bouguen, Adrien

Publication Date
2022-04-27

DOI
10.26085/C3588M

Series Name: WPS
Paper No.: 088
Issue Date: 27 April 2022

Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds, Conditional LATE and Machine Learning

Jan Berkes and Adrien Bouguen

Working Paper Series

Center for Effective Global Action
University of California

Recommended Citation:

Berkes, Jan and Adrien Bouguen (2022): Estimating Preschool Impacts When Counterfactual Enrollment Varies: Bounds, Conditional LATE And Machine Learning. CEGA Working Paper Series No. WPS-88. Center for Effective Global Action. University of California, Berkeley. Text. https://doi.org/10.26085/C3588M

Estimating Preschool Impacts with close substitute : Bounds, conditional LATE and Predicted sub-LATE

Jan Berkes,* Adrien Bouguen†

January, 2022

Abstract

We study the impacts of a large preschool construction program where newly built preschools compete with lower quality existing preschools as well as home care. In this context, we highlight that 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 expected larger gains among the latter. Using data from an experiment conducted in Cambodia, we implement several empirical techniques to isolate the impact on children who would have stayed at home had they not been enrolled in the newly built preschools. We argue that the impact on these children is a central parameter in the preschool literature. We first implement a bounding approach to show that, under reasonable assumptions, the effect on children who would have stayed at home absent the program is high and significant (between 0.14 and 0.45 SD). We then implement two other empirical approaches (conditional LATE and predicted subLATEs) to pinpoint the effect on these children. We find consistent evidence that the impact on these children is large and significant (0.15 SD - 0.3 SD) while the effect on children who would have enrolled in a preexisting preschool (absent the newly built school) is small and insignificant.

JEL classification: I24, I25, J24

Keywords: Early Childcare Development (ECD), Education supply, Preschool, Close Substitutes, sub-LATEs.

Thanks to the World Bank and Deon Filmer, Tsuyoshi Fukao, Simeth Beng, and Samuel Fishman for their constant support; The SIEF team and Alaka Holla for their financial support; the Ministry of Education in Cambodia, and specifically Sok Sokhom and Lynn Dudley, who made this research possible; and Angkor Research and John Nicewinter, Ian Ramage, Benjamin Lamberet, and Kimhorth Keo, for the stellar fieldwork. 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, Alain de Janvry, Clément de Chaisemartin, Antoine Camous, Harald Fadinger, Luc Behaghel and Marc Gurgand.

*PhD candidate, DIW Berlin, Freie Universität Berlin

†Assistant Professor, Santa Clara University, California. Adrien Bouguen also acknowledges financial support from the German Research Foundation (DFG) project SFB 884 during his stay at the University of Mannheim

1 Introduction

Development programs, such as large infrastructure plans, new financial institutions, or new technologies, are often introduced in a context where access to similar services already exists. In program evaluation, the presence of these close substitute programs generate a variation in the counterfactual enrollment likely to affect the interpretation of the standard treatment effect parameters – intention-to-treat (ITT) and local average treatment effect (LATE). Specifically in this context, standard treatment effect parameters cloud the treatment effect differences between individuals who would have benefited from a close substitute program and those who would not (Heckman et al., 2000; Kline and Walters, 2016; Dean and Jayachandran, 2020). While standard treatment effect parameters remain internally valid and relevant estimates of the overall policy impact, clarifying how the effect of a policy depends on close substitutes is critical for producing evidence that is comparable across studies and for making appropriate policy recommendations.

In this paper, we estimate the impact of a preschool program in a Cambodian context, where newly built formal preschools (or community preschools) compete with existing alternative childcare arrangements. In this program, the construction of the community preschools was randomly assigned among villages with alternative forms of preschool (or alternative preschool). The study, therefore, creates two sub-populations of compliers: children who would have stayed in home care in the absence of the program (or *home compliers*) and children who would have attended alternative preschools (or *alternative compliers*). Consequently, the ITT effects reported in this paper measure the effectiveness of the new community preschools in comparison to a mix of home care and alternative preschools. In this paper, we propose to go beyond standard treatment effect parameters and develop strategies to isolate the specific contribution of the home compliers and the alternative compliers. Specifically, we will suggest that isolating the effect on home compliers is of prime importance for the early childcare development (ECD) literature.

The presence of close substitute programs is not a unique characteristic of our study. To our knowledge, every large-scale randomized controlled trial conducted to measure preschool effects in a low-income country is implemented in an environment where 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 attend preschool (Martinez et al., 2017), 16% do 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 articles present 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. Consequently, while Martinez et al. (2017) find strong effects of preschool attendance on child outcomes, Bouguen et al. (2018), Bouguen et al. (2013), Blimpo and Pugatch (2017), and Brinkman et al. (2017) in low-income countries, and Puma et al. (2012) in the US, find no effects or only small effects. We interpret this lack of consistency in the literature, at least partially, as a result of the specific substitution patterns that affect every preschool study.

Using new empirical strategies and detailed information about the alternative forms of preschool available to parents, we isolate the impact of the program on home compliers from the impact on alternative compliers. We argue that the impact on home compliers is a critical parameter in the (ECD) literature and that failure to isolate both sub-treatment effect parameters contributes to the ongoing confusion in the debate about the effectiveness of ECD in low-income countries.

We start our analysis by reporting the first year reduced-form estimates provided in our companion paper (Berkes et al., 2019).[1] The ITT effect on three- to five-year-old children varies from 0.046 to 0.061 standard deviations (SD) on a large set of child development measures (executive function, language, numeracy, fine-motor, and socio-emotional development) or 0.051 SD when we aggregate these tests in a *cognitive development index*. We provide more ITT results (using anthropometrics, parental measurements and effects after two years) in our companion paper Berkes et al. (2019).

We then document a large degree of program substitution, using detailed information about the alternative forms of preschool available to parents. The construction of the CPS causes 39% of the treatment group to enroll in CPS (0% in the control group as no CPSs were built in these villages) but we also observe an important reduction in APS enrollment (-28 pp). These results indicate that in the absence of the construction program, many children would have enrolled in other preschool programs. Hence, the reduced-form effects reflect both the treatment effect of CPS attendance on children who would have stayed at home – the effect on home compliers – but also the effect from enrolling in CPS instead of enrolling in another existing preschool or alternative preschool (APS) – the effect on alternative compliers. Since we cannot distinguish the alternative compliers from home complier, the sub-LATE parameters are not identified. However, we show that the share of a-compliers is known : it corresponds to the proportion of switchers (28 pp) divided by the total number of compliers (38 pp), hence 73 %.

We then show that, under plausible assumptions, the effect on home compliers can be bounded between the traditional local average treatment effect (LATE) of going to CPS and the LATE of going to any preschool. With these bounds, the effect

[1] As shown in (Berkes et al., 2019), the 2-years ITT effects are only significant for a sub-sample of children. Estimating the LATE and sub-LATEs at 2-year would therefore be mostly insignificant. We therefore restrict our analysis to the first year in this paper

on home compliers, who attended the new CPS program for about 9 months, varies between 0.13 SD and 0.45 SD (on our cognitive index) or between 0.14 SD and 0.39 SD when we use additional baseline variables to narrow our bounds. Finally, we use an empirical technique previously applied elsewhere, the conditional LATE (Kline and Walters, 2016; Hull, 2018), to obtain point estimates of the effects on home and alternative compliers. We find consistent evidence that the effect on home compliers is around 0.16 SD on a child development aggregate score. The effect on alternative compliers is positive but smaller and indistinguishable from zero. While consistent, these results rely on a heavy constant treatment assumption and are therefore fairly sensitive to the choice of instruments.

To go beyond the limitations of the bounds and of the conditional LATE, we introduce a new approach based on predicting, using our rich baseline sample and a LASSO algorithm, who is most likely to be a- and h-compliers within the group of children enrolled in CPS. We call this approach the predicted sub-LATEs. To predict who is most likely to be a a-complier, we first calculate, using baseline characteristics selected by a LASSO, the predicted probability to be enrolled in an APS at follow-up. We then apply this prediction to the children enrolled in CPS and consider as a-compliers the known share of a-compliers (here 73%) with the highest APS enrollment prediction. Using this approach, we find that the effect of home compliers to be larger than the one found with the conditional LATE approach (+0.27 SD) but still very consistent with the bounds. Here again, the effect on alternative compliers is estimated to be small and below our detection power (+0.055 SD). our results suggest that fairly small ITT results are entirely compatible with mid-to-large impacts on students who would have stayed otherwise, with +0.27 SD on the index of cognitive development representing two-third of the initial cognitive gap between children from relatively poorer and wealthier background.

Our article directly relates to the strand of applied literature that discusses the interpretation of treatment parameters in the presence of close substitutes (Heckman et al., 2000; Feller et al., 2016; Kline and Walters, 2016; Hull, 2018; Kirkeboen et al., 2016). As described by Kline and Walters (2016), in the preschool context, the local average treatment effect is a weighted average of the effects on home and alternative compliers. Yet, these *sub-LATEs* parameters cannot directly be derived, as the counterfactual care arrangement is not observed for individual children in the treatment group.

Depending on the objectives of the researcher, the identification of sub-LATEs might not be of prime concern. As noted by Kline and Walters (2016), program substitution can even be seen as an opportunity when estimating the cost-effectiveness of a similar policy. When the substitution patterns replicate those that would have been found in an ecological environment, then ITT and the standard LATE are the policy relevant parameters. Failure to isolate sub-LATEs and, in particular, failure to isolate the effect on home compliers is, nevertheless, an important limitation. First, the justification for ECD interventions relies heavily on the idea that formalized ECD programs should compensate unfavorable early environments at home (Cunha et al., 2010; Heckman, 2010; Cunha and Heckman, 2007; Cunha et al., 2013). This idea prevails in the United States (Campbell et al., 2002; Currie, 2001; Heckman et al., 2010) and in low-income countries (Gertler et al., 2014; Walker et al., 2011). Preschool interventions, nutrition supplementation, and cognitive stimulation programs for children aged 0–6 are usually seen as ways to compensate for detrimental factors in the home environment. Failure to isolate the benefit of preschool versus home environment is very detrimental to the ECD literature.

Second, many influential empirical papers in the early childcare literature implicitly report the impacts on the home compliers. The Jamaica study (Grantham-McGregor et al., 1991) in low-income countries and the Perry Preschool Project (Anderson, 2008) in the US, which constitute the empirical foundation for new ECD interventions, implicitly measure effects on home compliers. Comparing more recent at-scale programs with these studies on the basis of reduced-form estimates is inappropriate if children in the control group have access to close substitutes.[2] More generally, since the magnitude of standard treatment parameters crucially depends on local conditions – including rate of substitution and substitute programs' quality – the ITT and LATE are likely to be systematically incomparable across contexts. Instead, the effect on home compliers does not depend on close substitutes and comparability can be assessed using commonly available socioeconomic characteristics, e.g. parental education, poverty, and stunting rates.[3] Any aggregate meta-statistic about the effectiveness of ECD interventions that does not take substitution into account is of limited value. With the increased concern around reproducibility and the revived interest around meta-analysis (Meager, 2018), we believe this is a crucial limitation.

Third, in order to make appropriate policy recommendations, understanding the expected substitution patterns and isolating the sub-LATEs is complementary to a reduced-form analysis. If the share of home compliers is small, or if the effect on alternative compliers is null or negative, then large treatment effects on home compliers are entirely consistent with, for instance, low and insignificant ITT effects. The sub-LATE analysis informs policymakers that the same program, targeting, for instance, home compliers, could generate substantial impacts. It could further mean that additional demand-side interventions (information, cash transfers, nudges, free lunch, free transportation) should be implemented to attract those children who

[2]In fact, when the counterfactual care arrangement is a close substitute for a majority of compliers, the study might be more comparable to quality interventions, such as the study by Ozler et al. (2018), which evaluates the impact of preschool quality improvement on child performance in Malawi and who implicitly measure an effect on alternative compliers.

[3]Similarly, although less easily observed, characteristics of the close substitute program can be used to assess the comparability of the $LATE_{ac}$.

would benefit most from the program.

This paper proceeds as follows. First, we describe institutional details, the experimental design, and the studied sample. Second, we present the empirical framework that is used to analyze the data. We focus on the relationship between ITT, LATE, and sub-LATE parameters. Third, we present our reduced form estimates: the adherence to the experimental protocol, the preschool participation, and the impact on children's performance after one year of preschool. In the fourth section, we present our estimations of the treatment effect on home compliers. We discuss the validity of our bounds and then provide one alternative strategies to point estimate the sub-LATEs.

2 Background, Data, and Design

2.1 Recent ECD Program Development in Cambodia

Despite robust economic growth since 2000, Cambodia remains one of the least developed countries in Southeast Asia, with a GDP per capita estimated at $1,384 in 2017 ($4,000 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), the country fares poorly in comparison to its neighbors, Thailand and Vietnam.[4] To increase the capacities and quality of its education system, the government of Cambodia, with the support of the World Bank, launched an education expansion program for the 2014—2018 period, called the Global Partnership for Education II (GPE II). Berkes et al. (2019) provides further details about previous education expansion programs (GPE I).[5] GPE II, and the education sector in Cambodia. This paper focuses on the part of the GPE II that includes the construction of community preschools (CPSs).

2.2 Formal Community and Alternative Preschool Programs

Before GPE II, two distinct types of public preschools existed in Cambodia: state preschools (SPS) and (informal) community preschools. Since community preschools lacked uniform quality standards, we refer to them as informal (community) preschools (IPSs). In this article, we consider both IPSs and SPSs as alternative preschools (APSs).[6,7] GPE II introduced a new type of community preschool with a uniform

[4]Source: Data from UNESCO Institute for Statistics.

[5]Bouguen et al. (2018) analyzed the impact of the preschool construction funded by GPE I.

[6]According to government data (MoEYS, 2017), out of 7,241 preschool facilities in Cambodia in 2016, 55% were SPSs, 39% were IPSs, and 6% were private preschools. However, these preschools are not evenly distributed across the country and 38% of the 1646 communes in Cambodia had no preschool facility.

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

quality standard, which we refer to as (formal) community preschool (CPS).

State preschools are financed by the Ministry of Education, Youth, and Sports (MoEYS) (see Figure 1 for pictures of a typical SPS facility). SPS teachers benefit from two years of formal training in a MoEYS teacher training center in Phnom Penh. They receive a monthly salary of about 180 $ in 2017 to teach for three hours a day, five days a week. As almost all SPSs are attached to a public primary school, SPSs have access to properly equipped classrooms, as well as teaching materials, play materials, and sanitary facilities.

In contrast, informal community preschools are not typically attached to a primary school. Local communities establish IPSs and cover operational costs. This includes the IPS teacher salary, which is at the discretion of the local commune council. It varies from $30 to $50 per month, with most IPS teachers relying on additional sources of income. IPS teachers are trained for about 35 days by provincial education departments before they begin working. Teachers are required to provide a 2-hour preschool class, five days a week. The quality of IPSs can differ substantially across villages as, until 2018, communes were required to establish IPSs using their own funds. Consequently, IPS classes are often held in a teacher's home, in a community hall, or a pagoda (see Figure 2). IPSs often lack appropriate equipment, such as teaching and play materials or sanitary facilities. In most cases, IPSs even lack the most rudimentary equipment, such as tables and chairs.

To increase preschool access and to improve the unsatisfactory quality of IPSs, the Cambodian government agreed to use the GPE II grant to establish 500 new formal community preschools. Some of these CPSs replaced existing informal arrangements; others were established in villages that had no previous preschool or were too large to be serviced by one preschool alone. Unlike IPSs, a CPS benefits from uniform quality standards, such as a standardized building (see Figure 3), directly financed by the GPE II. CPSs have a capacity of 25 children and are fully equipped with tables, chairs, a blackboard, and teaching materials. In partnership with GPE representatives, MoEYS is responsible for the curriculum, teacher recruitment, and teacher training, as well as the monitoring of the running facility, including regular payment of teacher salaries. The CPS teacher is usually a (female) community member who receives training from the ministry and gives a two-hour class each day, five days a week, to children aged three to five years. Importantly, CPSs, SPSs and IPSs are officially costless for parents.[8]

[8]In some villages as it appeared, the village chief asked families to make a contribute to improve the preschool. For instance, CPSs did not include a latrine and we know that some villages built a latrine next to the school with funding coming from the community. We do believe that these were voluntary donations though. When asked why parents did not send their child to school, only 2% of them responded for "financial reasons".

2.3 Randomization and Data

The evaluation of the CPS program is based on a cluster randomized controlled trial.[9] All sample villages are situated in the south and northeast parts of Cambodia, as the western part of the country had already been covered by previous expansion plan. Eligibility criteria for villages to participate in the study were demand for a CPS, a high poverty rate, and a high number of children between the ages of 0 and 5.

The total study sample comprises 305 villages. Before baseline, we randomly assigned these villages to different treatment branches: a control group (58 villages), which received no GPE II intervention; and a CPS treatment group (120 villages), which received a CPS. An additional 127 villages received a CPS plus a demand-side intervention.[10] These demand-side interventions were in part implemented during follow-up data collection in 2017 and, hence, their impact is evaluated on the basis of a follow-up in 2018; this is the focus of a separate article (Berkes et al., 2019).

Table 1 provides an overview of data collection activities and the timing of preschool construction. The analyses presented in this paper are based on two main waves of data collection: a baseline data collection in 2016 and an initial follow-up in 2017. Additionally, a brief monitoring survey was conducted in late 2016 to confirm that preschool construction proceeded as scheduled. With 86% of CPS constructed before follow-up, Table 1 confirms that the construction plan was almost perfectly respected. Yet, despite our effort to conjointly deploy the preschool construction and baseline survey, in 17% of the treatment group villages, the CPS was already available at baseline. Conducting a social experiment on school construction is challenging, since conducting baseline too early (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.[11] Inversely, conducting the baseline too late would have resulted in baseline measures that are already affected by the program. In Section 2.4, we discuss the implications of the slight overlap between the baseline survey and construction.

During the baseline data collection exercise in 2016, our survey firm sampled up to 26 eligible households per village.[12] Eligible households are composed of at least

[9] The study was pre-registered at the AEA's Social Science Registry (AEARCTR-0001045).

[10] We randomly assigned the remaining 127 villages to two variants of the demand-side interventions (an awareness campaign or an awareness campaign plus a parenting program) to stimulate preschool enrollment. We performed the randomization with province-level stratification on a list of 310 eligible villages provided by MoEYS. Of these, 60 were assigned to the control group, 123 to T1, 63 to T2, and 64 to T3. Unfortunately, the list contained erroneous village names, with 5 either duplicated or unable to be identified following randomization. Therefore, the total number of villages decreased to 305. We treated this drop-out as random and did not replace the villages.

[11] As described in Bouguen et al. (2013), construction delays occurred in a previously evaluated program in Cambodia, which considerably reduced take-up and statistical precision.

[12] They used an adapted version of the EPI walk to sample the household. EPI refers to the Expanded Programme on Immunization of the World Health Organization; see e.g. Henderson and Sundaresan (1982).

one child between 24 and 59 months old at baseline. Thus, eligible children were between three to five years old at follow-up.

Our survey instruments include a village, teacher, household, and caregiver survey, as well as a child assessment.[13] The village and teacher surveys serve as sources of information about village and preschool infrastructure. The household survey captures information about household wealth, income, and other socioeconomic measures. The caregiver survey is used to obtain information about parenting practices, a fluid intelligence measure of the caregiver (based on Raven's Progressive Matrices), and detailed information about the child (for example, preschool enrollment history). Parental-reported versions of the Strengths and Difficulties Questionnaire (SDQ) and the social development scale of the Malawi Development Assessment Tool (MDAT) were administered to caregivers to obtain a measure of socio-emotional development of the children. Additionally, a comprehensive child assessment was conducted. The battery of child tests measure five crucial domains of cognitive and physical child development: executive function, language, numeracy, as well as both fine- and gross-motor development.[14] Most child tests stem from the Measuring Early Learning Quality and Outcomes project (MELQO). MELQO tools are designed to provide a starting point for national-level adaptation of global measures of child development (see UNESCO (2017) for an overview) and demonstrate adequate internal validity (Fernald et al., 2017; Berkes et al., 2019).[15] Additionally, anthropometric measurements (height and weight) are taken from all tested children.

2.4 Sample Description and Cognitive Inequality

A summary of the study sample is presented in Table 2. The baseline sample includes 4075 households and 4393 children aged between 2 and 4 in 178 villages.[16] For 4315 out of 4393 children, consent to participate in the child assessments was obtained from the caregivers and children.[17] Table 2 also gives an overview of the households interviewed at the follow-up in 2017. The attrition rate, 8.8% for household attrition, can almost entirely be explained by seasonal or permanent relocation of households, since the study does not follow up on households that move beyond the boundaries of the sample villages. Attrition is slightly larger for

[13]The caregiver is defined as the direct relative (parent, grandparent, aunt/uncle, or adult sibling) who takes care of the child most of the time. In most cases, the caregiver is a biological parent (60.4% at baseline, 58.7% at follow-up). In the provinces of Kampong Speu, Kandal, Prey Veng, Svay Rieng, and Takeo, the caregiver is often a grandparent. These are provinces with relatively high levels of manufacturing and mothers are frequently absent during the day.

[14]We discuss cultural adaptation, content, and scoring of all child test scores and the parental practices measures at length in Berkes et al. (2019).

[15]Our version of the test is available upon request.

[16]Unless otherwise stated, all numbers in this paper refer to the sample of 178 villages without the additional treatment groups. On the full sample of 305 villages, the sample includes 7053 eligible households and 7546 children between 2 and 4 years of age.

[17]The 78 eligible children without baseline test scores are balanced across treatment and control (2% versus 1.67%).

children (10.6%) but the difference of attrition between the treatment and control group remains small (1.9%) and insignificant.

Table A1 (household and caregiver characteristics) and Table A2 (child characteristics) show a balance in variables between treatment and control group separately for the baseline sample and the sample of households who participated in baseline and follow-up. The tables show that the variables are balanced at baseline and remain balanced after taking attrition into account (*Baseline and Follow-up Sample* panel). One exception is preschool enrollment caused by the slight overlap of preschool construction and timing of the baseline survey (cf. Table 1). Treatment children were 6.6 pp more likely to be enrolled in preschool (last panel, Table A2). As discussed before and indicated in Table 1, the difference is due to the fact that in 17% of treatment villages, the CPS was completed briefly before the baseline survey. Since the treatment children only spent 11 more days in preschool than the control children and since we do not measure any developmental difference between treatment and control at baseline, we consider the difference as negligible.

Table A1 shows variables that characterize the socioeconomic background of our sample population. Households are generally poor – 41% are considered as poor, according to our multidimensional poverty index.[18] In our sample, 55% of households live on less than $100 per month. Caregivers, on average, have years of formal education. Based on WHO Child Growth Standards, 34% of tested children are stunted and 10% suffer from wasting.

Child test scores are strongly associated with socioeconomic background characteristics. As described more at length in Berkes et al. (2019), children in the top wealth quintile perform, on average, between 0.46 and 0.7 SD better than children in the bottom quintile. Schady et al. (2015) find similar results in South America. The gap that separates children age 3–5 from the bottom quintile and the top quintile corresponds to about 6–12 months of cognitive development. Thus, wealthier children are up to one year ahead of poorer children in development once they reach primary school age.

2.5 Preschool Quality

We use village survey data to show differences in quality measures between the types of preschools at baseline (Table A3) and follow-up (Table A4).[19] These simple

[18]We construct a binary poverty index using baseline data and an adapted version of the method by (Alkire and Santos, 2010). A household is considered poor if it is deprived in at least 30 percent of the weighted indicators for health, education, and living standards.

[19]Note that the full study sample of all 305 villages is used in these tables to maximize statistical power. Since CPSs were also constructed in the two other treatment branches and since both SPSs and IPSs are present, they can be used to document preschool quality. Further, note that, as shown in Table 1, only a handful of CPSs was already open at baseline, while almost all CPSs were completed at follow-up. Hence, Table A4 is better suited to assess the final quality of CPSS.

comparisons should be analyzed carefully as they likely suffer from selection bias.[20] Table A3 documents that SPSs are significantly different from CPSs and IPSs. SPSs are larger (6 additional children when compared to an IPS, which serves around 21 students) and they have more equipment, such as chairs, tables, and blackboards. SPSs also have fewer significant problems, as reported by the village chief. SPS teachers benefited from more training days (+152 days, or about three times as much) and they are more likely to be paid regularly with a significantly higher salary (on average, \$90 per month versus \$35 for IPS teachers). Already at baseline, the quality of CPSs appears better than that of IPSs:. CPSs have better, more spacious, buildings and enjoy more resources. In addition, their teachers were also paid more regularly. Yet village chiefs considered CPS and IPS teachers as comparable in terms of salary and training.

At follow-up (Table A4), SPSs still offered a higher quality than CPSs and IPSs, but CPSs quality had further increased. CPS buildings are still reported to be larger and of better quality than IPSs, but this time, CPSs are reported to have more tables, chairs, and additional learning materials. Indeed, at the time of the follow-up survey, almost all CPS equipment had been delivered. Yet, again, in terms of teacher quality, the difference between IPSs and CPSs is small, at least in the eyes of the village chief. Teachers in IPSs and CPSs seem to have seen their situations improve in similar fashions: preschool teachers are more regularly and better paid at follow-up than they were at baseline. Additional information on the difference between IPSs, SPSs, and CPSs, relying on in-class observations and additional follow-up surveys, are available in Berkes et al. (2019). While Berkes et al. (2019) confirm that IPS and CPS teachers share many characteristics (age, gender, education, experience), the equipment, the class setting (time hours effectively teaching), as well as the educational content (following curriculum...) is reported to be significantly superior in CPSs than in IPSs. SPSs remain superior to CPSs and IPSs across all measured dimensions. In all, Tables A3 and A4 indicate that the program has significantly improved the infrastructure quality of the community preschools: CPSs have more materials and better premises. Yet the teaching quality – arguably the most important factor in early childhood development – remains comparable in the eyes of village chiefs.

3 Empirical Framework

As explained in the previous section, we evaluate the impact of the CPSs in the context where alternative preschools (APSs), i.e. SPSs and IPSs, are also available. In this section, we outline the empirical framework and list the strategies we im-

[20] For instance, at baseline (Table A3) only a handful of CPSs was constructed and they may be of different quality than the average CPSs. Similarly, at follow-up (Table A4, many IPSs have closed and were replaced by CPSs and the remaining IPSs may be different from the average one.

plement to identify the relevant treatment parameters. We extensively discuss the identification strategies and their assumptions in the Appendix.

3.1 Identifying CPS Impacts using Traditional Parameters

To estimate the intention-to-treat effect (ITT) and the local average treatment effect (LATE) of attending CPS, here called $LATE_{cps}$, we define Z_i to be the instrumental variable that takes the value 1 for children in treatment villages and 0 otherwise. Further, we define D the participation variable that takes value c, a or h depending on whether the child is enrolled in CPS (c), in APS (a) or is staying at school (h). As described in the Appendix, the ITT is valid under Assumption A1 (independence) and A2 (SUTVA), which are very likely valid in our setting. We estimate ITT effects using the following regression:

$$Y_{iv} = \beta_0^{ITT} + \beta_1^{ITT} Z_v + \boldsymbol{W_{iv}} \boldsymbol{\beta_2^{ITT}} + \mu_v^{ITT} + \epsilon_{iv}^{ITT} \tag{1}$$

where Y_{iv} denotes the *observed* outcome of child i and village v, Z_v the *observed* treatment assignment, and $\boldsymbol{W_{iv}}$ a set of control variables. μ_v and ε_{iv} are the village-specific error term component and the unobserved within-village error component, both assumed to be uncorrelated with $\boldsymbol{W}$ and Z. We use standard errors clustered at the village level to account for the randomization implemented at the village level. The outcomes of interest Y include (i) the school construction collected at the village level (see supra Section 4.1); (ii) the enrollment in, and months of exposure to, each childcare arrangement (see supra Section 4.1); and (iii) the children's cognitive development (see supra Section 4.2).[21] We estimate equation (1) using two sets of control variables, one restricted to stratification variables (province fixed effect), the second additionally includes baseline test scores, as well as the age dummies and gender.[22] Since the latter specification considerably reduces the residual variation (R^2 well above 50%) and was pre-announced in our pre-analysis plan, this is our preferred estimation.

We estimate the local average treatment effect CPS ($LATE_{cps}$) using following equation:

$$Y_{iv} = \beta_0^{LATE_{cps}} + \beta_1^{LATE_{cps}} \mathbb{1}_{\{D_i=c\}} + \boldsymbol{W_{iv}} \boldsymbol{\beta_2^{LATE_{cps}}} + \mu_v^{LATE_{cps}} + \epsilon_{iv}^{LATE_{cps}} \tag{2}$$

[21]Additional outcomes are presented in (Berkes et al., 2019): anthropometrics measures, parental response to school construction and socio-emotional development. In this companion paper we also analyse impacts two years after the beginning of the program.

[22]We replace missing baseline test scores with the sample mean. We create a dichotomous variable indicating when a missing value was imputed. We add these missing value dummies to the regression in the second specification. Age is measured as a trimester fixed effect and also imputed if missing.

where $\mathbb{1}_{\{D_i=c\}}$, which takes value 1 when the child go to CPS (0 otherwise), is instrumented using Z_v. As shown in Appendix 5, the LATE CPS is valid under additional assumption A3 (non-zero average causal effect), A4 (extended exclusion restriction) and A5 (extended monotonicity assumption). The LATE estimation is valid, yet, due to the different counterfactuals in this context, there are implicit assumptions behind this standard model which we, following the framework by Kline and Walters (2016), discuss in Section 3.2. The main takeaway is that by defining $D_i \in \{c, a, h\}$ as capturing enrollment in CPS, into APS, or as not being enrolled in preschool (home care), two different types of compliers and never-takers exist:

1. a-never takers (ANT): $D_i(0) = a$, $D_i(1) = a$,
2. h-never takers (HNT): $D_i(0) = h$, $D_i(1) = h$,
3. a-compliers (AC): $D_i(0) = a$, $D_i(1) = c$,
4. h-compliers (HC): $D_i(0) = h$, $D_i(1) = c$,

In Figure 4 we provide a visual representation of the different sub-populations in our sample under the stated assumptions. In the absence of the intervention, i.e. before the CPS construction, children either stay home or attend an APS. After construction, the presence of APSs generates two types of compliers, the children who would have enrolled in an APS absent the program and the children who would have stayed at home. While the $LATE_{cps}$ is identified using Z as an instrument, the respective sub-LATEs, $LATE_{hc}$ and $LATE_{ac}$ are not because we do not know who, among the goupe of compliers, is h- or a-type. After describing these two populations of compliers, We present below three strategies to identify them.

3.2 Substitution and sub-LATEs

Under the defined assumptions (A1–A5), the local average treatment effect of going to CPS, here called $LATE_{cps}$, is identified and given by:

$$LATE_{cps} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i = c|Z_i = 1] - E[D_i = c|Z_i = 0]}.$$

In presence of a close substitute, it can be shown that the $LATE_{cps}$ can be further decomposed into of the two sub-LATEs (Kline and Walters, 2016):

$$LATE_{cps} = S_{ac}LATE_{ac} + (1 - S_{ac})LATE_{hc} \tag{3}$$

where $LATE_{ac} \equiv E[Y_i(c) - Y_i(a)|D_i(1) = c, D_i(0) = a]$ and $LATE_{hc} \equiv E[Y_i(c) - Y_i(h)|D_i(1) = c, D_i(0) = h]$ give the average treatment effect on a and h compliers. Importantly, S_{ac}, the share of a-compliers (within the group of compliers), is identified and given by Kline and Walters (2016):

$$S_{ac} = \frac{P(D = a|Z = 0) - P(D = a|Z = 1)}{P(D = c|Z = 1) - P(D = c|Z = 0)} \tag{4}$$

Figure 4 provides a visual representation of the parameters in Equation (3): 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 $LATE_{hc}$.

3.3 Identifying Bounds for $LATE_{hc}$

Equation (3) makes explicit the challenges faced by researchers when estimating the impact of a policy in a context of close substitutes. Under A1–A5 (see Appendix), $LATE_{cps}$ is identified, but its sub-LATE components ($LATE_{ac}$ and $LATE_{hc}$) are not. Yet, under plausible assumptions, we can derived bounds for the $LATE_{hc}$ a parameter of particular interest for the preschool literature. The assumption to derive the bounds is:

$$0 \leq LATE_{ac} \leq LATE_{hc} \tag{5}$$

i.e., the CPSs offer, on average, a better learning environment than APSs (left hand side of the inequality) and that the returns to CPS are not higher for a-compliers (right hand side of the inequality).

The left hand side of inequality (5), $0 \leq LATE_{ac}$, simply implies that switching from an APS to a CPS is not, on average, detrimental to a-compliers. Given the resources devoted to CPSs in comparison with IPSs, as discussed in Section 2.5, we believe that the left hand side of (5) inequality is a very likely assumption. The right side of inequality ((5)), $LATE_{ac} \leq LATE_{hc}$, implies that a-compliers do not benefit more from the CPS than the h-compliers.There are reasons to believe that this is also a likely assumption. Intuitively, h-compliers benefit from a larger improvement of their learning environment than a-compliers, who already benefit from some preschool intervention regardless of their treatment status. Consequently, the h-compliers will likely benefit at least as much from CPS than will the a-compliers.

Under (5), we can calculate a lower bound (LB) and upper bound (UB) to $LATE_{hc}$ with:

$$LATE_{ac} = LATE_{hc} \iff LATE_{hc}^{LB} = LATE_{cps} \tag{6}$$

The low bound assumes that h-compliers and a-compliers equally benefit, on average, from the CPS intervention . Hence, $LATE_{cps}$ is the local average treatment effect of both sub-populations. Under the upper bound, we assume:

$$
\begin{aligned}
LATE_{ac} = 0 \iff LATE_{hc}^{UB} =& \frac{LATE_{cps}}{(1 - S_{ac})} \\
=& \frac{\beta_1^{ITT}}{\beta_1^{FS_c} * (1 - S_{ac})} = \frac{\beta_1^{ITT}}{\beta_1^{FS_c} * (1 + \frac{\beta_1^{FS_a}}{\beta_1^{FS_c}})} \\
=& \frac{\beta_1^{ITT}}{\beta_1^{FS_c} + \beta_1^{FS_a}} = \frac{\beta_1^{ITT}}{\beta_1^{FS_{ps}}} \equiv LATE_{ps}
\end{aligned} \tag{7}
$$

with $LATE_{cps} = \frac{\beta_1^{ITT}}{\beta_1^{FSc}}$, β_1^{FSc} the CPS first stage parameter described in Appendix equation 14, β_1^{FSa} the equivalent first stage parameter for the APSs, $\beta_1^{FS_{ps}} = \beta_1^{FSa} + \beta_1^{FSc}$ 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 $LATE_{ps}$.

3.4 Narrowing the Bounds for $LATE_{hc}$

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

$$
0 \leq LATE_{ac}(\boldsymbol{B}) \leq LATE_{hc}(\boldsymbol{B}) \tag{8}
$$

which is the equivalent to equation (8) for each value of $\boldsymbol{B}$, the variables orthogonal to Z. We can implement the bounding strategy – calculate $LATE_{cps}$ and $LATE_{ps}$ for each value of $\boldsymbol{B}$ – in the sample cells formed by $\boldsymbol{B}$ if B is categorical. We then average across the value of $\boldsymbol{B}$ to recover the unconditional narrow lower and upper bounds, using the probability to belong to the cells as weights. The lower bound can also be more flexibly estimated using the following IV regression model:[23]

$$
Y_{iv} = \beta_0 + \beta_1^{lb} \mathbb{1}_{\{D_i = c\}} + \boldsymbol{B}_i \boldsymbol{\beta_2^{lb}} + \boldsymbol{W}_i \boldsymbol{\beta_3^{lb}} + u_{iv}
$$

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

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

bounds. Similarly, the narrow upper bound is estimated using :

$$Y_{iv} = \beta_0 + \beta_1^{ub} \mathbb{1}_{\{D_i=a \,\cup\, D_i=c\}} + \boldsymbol{B}_i \boldsymbol{\beta_2^{ub}} + \boldsymbol{W}_i \boldsymbol{\beta_3^{ub}} + u_{iv} \tag{9}$$

where $\mathbb{1}_{\{D_i=a \,\cup\, D_i=c\}}$ – the dummy for any preschool enrollment (CPS or APS) – is instrumented by $\boldsymbol{B}$, Z, $\boldsymbol{B} * Z$ and controlling for $\boldsymbol{W}$. While any baseline variable can theoretically be included in $\boldsymbol{B}$, the choice of the B variables depends on two potentially conflicting criteria. First, the size of the narrow bounds will depend on the ability of the $\boldsymbol{B}$ variables to predict the enrollment behavior. Second, the $\boldsymbol{B}$ variables should be sufficiently parsimonious to maintain a reasonable sample size in each cell (and therefore statistical precision). Also, assumption (8) needs to hold in each cell forms by $\boldsymbol{B}$, an additional argument to limit the number of cells form by $\boldsymbol{B}$. To balance both criteria, we estimate the narrow bounds using the median[24] of the APS enrollment's predicted value as defined below in equation (12), which allows both a reasonable sample size in both cells and a good prediction of the enrollment behavior. We show that using the median of the APS enrollment's predicted value significantly narrows the bounds.

3.5 Estimating the sub-LATE using Conditional LATE

Although the bounds rely on a less restrictive set of assumptions, we implement an alternative strategy to identify more precisely the sub-LATE parameters. Using additional baseline characteristics as instruments and under a constant treatment assumption, we can isolate $LATE_{hc}$ and $LATE_{ac}$. Kline and Walters (2016) – see also Hull (2018) and Feller et al. (2016) – show that sub-LATEs can be identified by interacting the randomly assigned preschool construction program with observed covariates. The structural equation takes the following form:

$$Y_{iv} = \beta_0 + \beta_1 \ \mathbb{1}_{\{D_i=c\}} + \beta_2 \mathbb{1}_{\{D_i=a\}} + \boldsymbol{X}_i \boldsymbol{\beta_3} + \boldsymbol{W}_i \boldsymbol{\beta_3} + u_{iv} \tag{10}$$

where Y_i is a follow-up outcome, $\boldsymbol{X}$ a set of additional variables orthogonal to the treatment, and $\boldsymbol{W}$ the preferred set of control variables used for the reduced form estimation. β_1 captures the $LATE_{hc}$, and β_2 captures the effect of going to APS. To derive the $LATE_{ac}$, we subtract β_2 from β_1. $\mathbb{1}_{\{D_i=c\}}$ and $\mathbb{1}_{\{D_i=a\}}$ are both endogenous and instrumented by:

$$\begin{aligned} \mathbb{1}_{\{D_i=c\}} &= \alpha_0 + \alpha_1 Z_v + \alpha_2 Z_v * \boldsymbol{X}_i + \boldsymbol{X}_i \boldsymbol{\alpha_3} + \boldsymbol{W}_i \boldsymbol{\alpha_4} + \mu_v + \epsilon_{iv} \\ \mathbb{1}_{\{D_i=a\}} &= \gamma_0 + \gamma_1 Z_v + \gamma_2 Z_v * \boldsymbol{X}_i + \boldsymbol{X}_i \boldsymbol{\gamma_3} + \boldsymbol{W}_i \boldsymbol{\gamma_4} + \phi_v + \nu_{iv} \end{aligned} \tag{11}$$

[24]we create a dummy taking value 1 when the child's APS enrollment predicted value is above the median

The identification of β_1 and β_2 relies on the independence of Z and $Z * \boldsymbol{X}$ and on the assumption that the h and a compliers have a constant return to preschool on $\boldsymbol{X}$. Importantly, the constant treatment effect assumption means that the X instruments should 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 APS) but not other forms of more standard heterogeneity (e.g. girls benefiting more from preschool or top performing children at baseline not benefiting as much as less advanced children). Even if the treatment effect varies with observed baseline characteristics, the new instruments (X) should only capture the heterogeneity caused by children's counterfactual enrollment. Arguably, variables at the village level that capture infrastructure quality are less likely to be correlated with the standard heterogeneity of *observed characteristics*. Inversely, variables at the children level are more subject to a violation of the constant treatment effect assumption. Note that when $\boldsymbol{X}$ is made of more than one instrument, the validity of this constant treatment assumption can be tested using an over-identification test.

3.6 Predicting a- and h-compliers and estimating sub-LATEs

Finally, we introduce a new strategy to estimate the sub-LATE that relies on predicting who are the a- and h- compliers using predicted probability of enrolling in APS at follow-up. We call this strategy the predicted sub-LATE approach. This approach relies on the fact that while we are unable to distinguish the a-compliers from the h-compliers within the group of students enrolled in CPS in the treatment group, we can estimate the *share* of a-compliers – S_{ac} see equation (4), in our case S_{ac}=73%. If we can predict who are the $100 * S_{ac}$% most likely to be enrolled in APSs in the group of children enrolled in CPSs in the treatment group, we can solved the identification problem. To determine who are the most likely to be a-compliers, we first predict the probability to be a child enrolled in an APS at follow-up. We estimate the following model:

$$\mathbb{1}_{D=a} = \kappa_0 + \boldsymbol{L_i \kappa_1} + \psi_v + \zeta_{iv} \tag{12}$$

using a logit model. We conduct this analysis in the control group to avoid capturing the treatment effect on a-compliers' status. We then calculate for each observation, the predicted probability to be enrolled in APSs at follow-up, $\hat{D_a}$. The choice of the $\boldsymbol{L}$ variables is crucial in this approach. We propose to select the best baseline variables to predict APS enrollment status at follow-up using a LASSO algorithm and 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 (1353 in our case). We exclude the text variables, the variables with very low response rate[25], and the variables without variation. We identify 217 eligible variables. We then create dummy indicators for categorical variables, we compute the square of each covariate to which we subtract all perfectly colinear variables.[26] We use the remaining 265 baseline variable to predict APS enrollment status at follow-up in the control group using a robust LASSO algorithm.[27] The robust LASSO algorithm identifies nine baseline variables that best predict follow-up APS enrollment in the control group: the age of the child (continuous), being four year-old at baseline (dummy), whether the child attend any school at baseline, whether the child attend preschool at baseline, the baseline index of all test scores , the baseline cognitive development index, the baseline anthropometrics index, household home equipment index and household wealth index. We use these nine variables as well as the province fixed effects[28] as $\mathbf{L}$ variables to estimate equation (12) using a logit regression model.

Finally, we use the predicted value $\hat{D_a}$ to rank CPS students in the treatment group from the highest probability to be a a-compliers to the lowest probability. We consider as a-compliers the $100 \cdot S_{ac}\%$ children with the highest $\hat{D_a}$ and respectively, as h-compliers the $100 \cdot (1 - S_{ac})\%$ children with the lowest $\hat{D_a}$. To estimate the a-complier sub-LATE, we then use the sub-sample composed of the students enrolled in APSs (in treatment and control) and the children predicted to be a-compliers to estimate via 2SLS the following regression:

$$Y_{iv} = \theta_0^a + \theta_1^a \mathbb{1}_{D=c} + \theta_3^a \boldsymbol{W_i} + \eta_{iv}^a \tag{13}$$

where $\mathbb{1}_{D_i=c}$ is instrumented by $\boldsymbol{Z}_v$ the treatment variable and $\boldsymbol{W}$ is the usual set of control variables. Figure 5 gives a visual representation of the approach. For the $LATE_{ac}$, we first identify the a-compliers children then estimate via 2SLS on the sub-sample made of a-compliers and children enrolled in APSs (bottom sample). Similarly, to identify the $LATE_{hc}$, we estimate the same model but on the sample composed of the children who stayed at home and the children identified as h-compliers in the procedure above.

The validity of the approach relies on the ability of the variables selected by the LASSO algorithm to separate, within the known group of compliers, the a-compliers from the h-compliers. One issue with this approach is misclassification of compliers within the group of students enrolled in CPS. Ideally the distribution of

[25]Typically less than 80% response at baseline among children tested at endline.

[26]To avoid dropping too many valuable observation, we impute baseline missing values using the average variable value. When we impute values, we create a dummy variable that takes value 1 for the imputed observation and 0 otherwise. We use these missing dummy as additional control variables.

[27]we partial out the LASSO with stratification variable i.e. province fixed effect and the treatment variable

[28]We also add the dummy variables that indicate which observations was imputed.

$\hat{D}_a$ is bimodal with a first peak, taking values close 0, representing $1 - S_{ac}\%$ of the distribution (here 27%) and a second peak, close to 1, representing $100 \cdot S_{ac}\%$. The size of the pseudo-R^2 when estimating equation (12) using a logit model may also provide indicative evidence of the validity of the approach.

4 Results

4.1 School Construction & Preschool Enrollment

We begin our empirical analysis by documenting in Table 3 ("construction") how treatment assignment affects preschool availability in the sampled villages. At follow-up, none of the control villages benefited from a CPS, while 86% of the treatment villages did. Given the constraint inherent to any construction work in low-income countries and the delays that such programs may incur, we consider this a particularly favorable result.[29] Yet, as noted, the construction occurs in the context where other preschools were available. The program therefore increases availability of any preschool by just 12 pp and causes many alternative preschools to shut down (-55 pp). This declines is essentially driven by IPSs while the availability of state preschools (SPSs) remains approximately unaffected (insignificant -6 pp). As confirmed by our field visits, IPSs were often shut or turned into CPSs as soon as the new preschool building became available. Conversely, SPSs, already a formalized form of preschool, remained available to the children. This substitution pattern has important implications for the interpretation of our results: Since IPSs are arguably of much lower quality than SPSs and CPSs (see Tables A3 and A4), the a-compliers are likely to contribute positively to the treatment effect. Although in the rest of the analysis, we will still consider the substitution pattern to exist between CPSs and APSs, the reader should keep in mind that the vast majority of the substitution is actually occurring between IPSs and CPSs.

To study the enrollment patterns at the child level, we explore the enrollment, separately by preschool type, in the second part of Table 3. Assignment to the treatment group significantly affects enrollment by about 39 pp. Since 14% of treatment group villages did not receive a CPS until follow-up (see Table 1), the CPS take-up rate in villages with CPS is about 45 pp on average, which is fairly high compared to the available literature.[30]

Finally, Table 3 provides child-level information about the substitution patterns.

[29] As a comparison, in Bouguen et al. (2018), differential construction rate was 43 pp. Martinez et al. (2017), and in Indonesia (Brinkman et al., 2017), all control villages received the program by follow-up. Blimpo and Pugatch (2017) do not provide information at the village level; however, compliance appears to be high in Mozambique (comparable to our setting) but lower in The Gambia.

[30] Martinez et al. (2017) report a differential take-up of about 33 pp, 24 pp in Indonesia (Brinkman et al., 2017), 9 pp in The Gambia (Jung and Hasan, 2016), and 2 5pp in the previous preschool impact evaluation in Cambodia (Bouguen et al., 2018).

As seen at the village level ("construction" panel), assignment to treatment negatively affects the probability to attend an APS ("enrollment" panel). This substitution is almost entirely driven by a CPS/IPS substitution. This confirms that the a-compliers are likely to substitute a poorly resourced preschool with the newly built CPS. Table 3 provides all the necessary information to calculate S_{ac} as in equation (4). Among the 38.9% of the treatment children who complied, 28.3% would have been enrolled in APS absent the CPS construction. They are a-compliers. The share of a-compliers is therefore 0.283/0.389=73% or 27% of the compliers would have stayed at home if the CPS construction had not occurred.

4.2 *ITT* Impacts on Children's Performance

We provide in Table 4 the first-year ITT impacts reported in our companion paper (Berkes et al., 2019)("ITT" columns). As indicated in Section 3.1, we present both the treatment-control differences controlling for province fixed effect in column 1 and the treatment coefficients controlled for baseline characteristics in column 2 (province fixed effect, gender, age and baseline test score). Given the high predictive power of the baseline variables (R^2 generally above 50%), the inclusion of control variables greatly reduces the standard errors. Since the PAP (Berkes et al., 2017) pre-defined the set of control[31] and the outcome[32] variables we use, column (2) is our preferred estimate.

ITT results point toward a positive effect of the CPS construction on the performance of children.[33] Children in treatment villages perform about 0.05 to 0.06 SD higher in treatment than those in control villages in cognitive test scores. The aggregation of each sub-test gives an overall cognitive impact of +0.051 SD for our preferred specification.[34] ITT results remain small in comparison with the ECD literature but, as mentioned, these are driven by both a- and h-compliers.

4.3 $LATE_{hc}$ Bounds and sub-LATEs Estimation

Although small, the magnitude of the ITT impacts reflects effects on both a- and h-compliers who may have benefited differently from the CPS. We start our inves-

[31]We have some minor deviation from the PAP: in the PAP, we loosely indicated province fixed effect, child, and household main characteristics, as well as baseline child performance measures. Our final set of controls include: child gender, child age (trimester fixed effects), province fixed effect, and all baseline child performance measures (test scores). Thus, we are more conservative than the PAP, as we do not include any household characteristics. Since household characteristics are very well balanced, none of them significantly improve, in term of precision, and none modified our results significantly.

[32]In the PAP, we included the gross and fine motor skills together. Given the low level of correlation between gross motor, and fine motor items, we decided to regroup fine motor skills with cognitive measures.

[33]Note that we also find impacts on parental involvement and perceived returns to education with parents spending more time in "cognitive games" and having a higher perceived return to primary and secondary education. We cover these impacts more in details in Berkes et al. (2019)

[34]Additional reduced form results are available in (Berkes et al., 2019)

tigation of the sub-LATEs by calculating the $LATE_{hc}$ bounds. We then use two strategies (conditional LATE and predicted LATE) to pinpoint the exact magnitude of the sub-LATEs.

4.3.1 Bounding $LATE_{hc}$

We start the investigation of the sub-treatment effects by estimating the bounds for the $LATE_{hc}$ in Table 4. As described in section 4.3.1, the lower bound is equivalent to estimating $LATE_{cps}$ and the upper bound is equivalent to $LATE_{ps}$. We start the analysis by looking at the effect on months of exposure, here measured as the number of months spent at (any) school since birth. Bounds for exposure indicate that the children who would have stayed at home (if the CPS had not been constructed) are now spending between three and nine months at school. For exposure, however, the $LATE_{hc}$ is probably closer to the upper bound. Indeed, children who switch from an APS to a CPS are *unlikely* to have experienced a different level of preschool exposure: we can assume that a-compliers spend about the same time at school irrespective of their treatment status. As a result, at least for exposure, we are more incline to assume $LATE_{ac} = 0$ (than $LATE_{ac}$=$LATE_{hc}$), which corresponds to the upper bound in Equation (7).[35] Consequently, the h-compliers are likely to have spent around 9 months at school in total. Interestingly, nine months is approximately consistent with the timeline of the impact evaluation (see Table 1) , but probably slightly too high.[36] Still, according to Table 4, the impact on the h-compliers is bounded between 0.13 SD and 0.45 SD for the cognitive development index and similarly for the individual scores.

We then estimate the narrow bounds using the median of the APS enrollment's predicted value ($\mathbb{1}_{\hat{D}_a>p50}$) as $\boldsymbol{B}$ variable. Narrow bounds presented in Table 4 are significantly tighter but are generally less significantly different from zero . The index of cognitive development $LATE_{hc}$ is predicted to be between 0.13 and 0.35 SD while other test scores vary between 0.13 and 0.4 SD, both consistent with mid-to-large treatment effects.

Overall, our results confirms that low ITT impacts are consistent with substantial effects on home compliers and that low ITT effects should not be interpreted as evidence against sizeable effects of preschool in general. In the following, we use additional identification strategies to obtain point estimates of the sub-LATE parameters.

[35] In theory, the exposure to IPSs and CPSs should be the same: lasts 9 months per year, 2 hours a day, 5 days a week. In practice, it is possible that some CPS opened later or that some IPSs closed more frequently during the school year. Yet, clearly, for exposure, $LATE_{ac}$ is closer to 0 than to the $LATE_{hc}$.

[36] The time that separated the school entry (October 2016) and the follow-up survey (April-May 2017) gives an exposure of 7-8 months. Yet, this is without considering that some CPSs did not open at the school entry but a little later. The exact average exposure is therefore likely to be slightly below 7.

4.3.2 Results from the Conditional LATE

Under the constant sub-LATE assumptions outlined in Section 3.5, we use additional instrumental variables interacted with Z to estimate the sub-LATE effects as in Equation (10). In Table 5, we provide a comparison between our bounds and the sub-LATEs for two sets of additional instruments. The first two columns provide the bounds, as in Table 4, with estimates for the $LATE_{hc}$ varying between 0.12 and 0.45 SD. We then estimate the sub-LATEs using, first, a set of village level characteristics (land area, presence of a primary school, presence of a secondary school and population size) and province fixed effects, and, second, our preferred estimation, where we add household and caregiver characteristics (caregiver education, caregiver IQ test, household poverty status) as instruments interacted with Z.[37]

Table 5 present several interesting results and one main caveat. First, our additional instruments have good first stage results: we reject under-identification in all cases with very low p-values. This suggests that the instruments we selected correctly predict APS and CPS enrollment. Second, in most cases, our instruments pass the over-identification, suggesting that these additional instruments properly capture the heterogeneity of counterfactual enrollment and that the constant treatment assumption is possibly valid. Third, our results for school exposure are very close to the upper bound. Since we anticipated the upper bound to be the valid assumption in Section 4.3.1, it gives credibility to the conditional LATE approach. If the conditional LATE was entirely inconsistent, the estimation for exposure had no reasons to be consistent with the upper bound.

Forth, our results of our preferred specification (last two columns) are fully consistent with our bounding exercise (and even with our narrow bounds). The $LATE_{hc}$ are generally close to the low bounds but positive and significant while the $LATE_{ac}$ are close to 0. This means that switching from APS to CPS did not improve children performance much while the overall treatment effect are driven by children who would have stayed at home if the CPS had not been constructed. Interestingly, the $LATE_{hc}$ are stronger and consistent with the bounds for Executive function, Language and Numeracy which can be considered as prime objectives for formalized preschool. The results are lower and not consistent with the bounds for socio-emotional skills and fine motor skills. For both outcomes, the $LATE_{ac}$ is estimated to be higher than the $LATE_{hc}$, a violation of the bounds assumption spelled out in equation (5). While this finding should be interpreted with care, as $LATE_{hc}$ and $LATE_{ac}$ are not significantly different from each other, the results suggest that APSs are particularly ineffective and potentially detrimental to the child development in term of socio-emotional and fine motor skills. This interpretation resonates with some of the existing empirical evidence on preschool impacts. Enrolling children

[37] When baseline variables are missing, we impute the missing values and create a dummy variable indicating which observation is imputed. We use these dummy variables as control in the regression.

in preschool too early is sometimes suspected to have negative effects on socio-emotional skills (Baker et al., 2008). Hence, it is possible, in our context, that APSs are so poorly equipped in material and building that they negatively affected socio-emotional performance. As a result, $LATE_{ac}$ (i.e. switching from APS to CPS) is large and (almost) significant, while better equipped preschools like CPSs, still positively improved cognition (+12 pp) as compared to home environment. Similarly, the $LATE_{ac}$ on fine motor skills is positive and significant (0.17 SD), while the effect on h- compliers is small and insignificant: this may also be interpreted as a limit of APSs. As shown in Tables A3 and A4, the main difference between IPSs and CPSs is in material and infrastructure: IPSs seems to lack the infrastructure necessary to allow children to develop their fine motor skills. For both skills, children may be better off staying at home than going to an APS. As a result, when a CPS is constructed, a-compliers benefit strongly from the intervention, while the effect on the h-compliers remains small and non-significant.

These results have one main caveat. The results are quite sensitive to the choice of $\boldsymbol{X}$ variable. For instance, results in the column "Prov.FE & village char." while not significantly different from our preferred specification, are not systematically consistent with the bounds. It is therefore possible that the conditional LATE is capturing other forms of heterogeneity which would cast doubt on its validity. Still the bounds, which do not rely on the constant treatment effect assumption, may serve as a benchmark to assess the validity of the conditional LATE.

Overall, although the estimated sub-LATEs are only valid under restrictive assumptions, both bounding strategies and conditional LATE converge to the same conclusion. $LATE_{hc}$ estimates are consistently closer to the low bound and entirely inconsistent with the upper bound. They are also, in most cases, consistent with our narrow bounds. Taken at face value, Table 5 suggests that for 8.5 months of exposure (about one school year), children who would have stayed at home in absence of CPS perform, on average, about 0.16 SD better on the summary index when enrolled in CPS. While moderately large, this effect corresponds to more than a third (38%) of the cognitive gap between relatively wealthier children (top wealth quintile) and relatively poorer children (bottom wealth quintile).

4.3.3 Predicted sub-LATE, an alternative to conditional LATE

Finally, we adopt another approach to identify the sub-LATE using the APS enrollment's prediction as in equation (12). As explained in section 3.6, we start by calculating the predicted value of enrolling in APS using the nine variables selected by the robust LASSO algorithm.[38] Using a logit model, we capture about 26 % of the APS enrollment variance. To verify the validly of the approach, we shows

[38] as well as the province fixed effect and for baseline variables with missing imputed values, the dummy indicating which observations have been imputed.

in Figure 6 the distribution of the predicted APS enrollment for students enrolled in CPS at follow-up. The distribution clearly shows two peaks as expected, one, smaller and close to 0, which regroups the children with a low probability to have registered in APS absent the CPS construction (h-compliers) and; a second, larger, which includes the a-compliers. In Figure 6, we also represent where the a- and h-compliers lies in the distribution with a-compliers constituting the $100 \cdot S_{ac}\%$, here 73% of the children enrolled in CPS with the highest $\hat{D}_a$. Respectively, the h-compliers are expected to be the $S_{ac}\%$ with the lowest $\hat{D}_a$. Figure 6 gives weight to the predicted LATE approach: the first peak represents more or less the 27% of CPS students who are likely to be h-compliers while the second peak is likely composed of a-compliers students. To be sure, between the two peaks, there are a number of children whose compliers status is less clear and who are likely to have been misclassified. Given that we only capture 26% of the APS enrollment variance, these misclassifications were to be expected and would constitute a violation to the validity of the approach. Yet, we tend to believe that this assumption is lighter than the constant treatment effect needed for the conditional LATE.

Assuming that the approach did classified a- and h-compliers correctly, we estimate the $LATE_{ac}$ via 2SLS using the sub-sample composed of students enrolled in APS and on CPS children predicted to have been enrolled in APS absent the CPS construction. We estimate the $LATE_{hc}$ using the same model on the sub-sample composed of students who stayed at home at follow-up and on CPS children predicted to have stayed at home absent the CPS construction (see Figure 5). We present our results in Table 6 together with results from the bounds and from the conditional LATE for comparison. We find some interesting results. First, the predicted LATE results on exposure deviate from the ones found previously. Using the predicted LATE, we find a $LATE_{hc}$ at +5 month of exposure smaller than our prior results. Since some CPSs opened after the school entry (in October 2016) and our follow-up measure was conducted in April 2017, an average exposure of 5 months is consistent however with the experiment timeline. Similarly, a $LATE_{ac}$ evaluated at 1.2 months may reflect the fact that CPSs are less likely to shut down or teacher less likely to be absent than in APSs.

Second, The $LATE_{hc}$ results are generally consistent with the bounds but systematically larger than the conditional LATE. On our main index of cognitive development, the predicted LATE approach estimate the $LATE_{hc}$ to be 0.27 SD, quite significantly higher than the one estimated with the conditional LATE and perfectly in line with the bounds. Consistently with the other approach, the $LATE_{ac}$ are rather small and not significantly different from zero. Using this method, we cannot confirm that the fine motor and socio-emotional development skills are differently affected than other dimension of cognition. The $LATE_{hc}$ socio-emotional skills is actually very high (0.46 SD) using this method.

Overall, our new approach to estimate sub-LATE are remarkably consistent with the two other approaches and confirm that the $LATE_{hc}$ is very probably large and significant in our context, corresponding to about two-third (64%) of the baseline cognitive gap between relatively poorer and relatively wealthier children.

5 Conclusion

In this article, we analyze the issue of close substitution and preschool impact in the context where other competing preschool programs (here called alternative preschools) are also available to parents and their children. We show that the presence of close substitute programs generates two different types of compliers: the children who would benefit from an alternative preschool (a-compliers) and those who would stay at home (h-compliers) in the absence of the program. Even though both groups of compliers may be similar in terms of observed characteristics, their local treatment effects are likely to be fundamentally different, because their counterfactual enrollment condition is different. Averaging together the treatment effects on both sub-populations, which is implicitly what standard treatment parameters (ITT and traditional LATE) produce, does not provide a sufficiently comprehensive picture of the way the program affects children's performance. We argue that, in addition to providing reduced-forms estimates, isolating the treatment effect on both sub-populations of compliers (sub-LATEs) is necessary to provide a comprehensive picture of the preschool impacts and make appropriate policy recommendations.

We rely on a large and well-implemented preschool construction program to produce three important results. First, we show that the preschool construction program increases preschool attendance (here by +39 pp) and slightly improves the reduced-form performance of three- to five-year-old children (+0.051 SD). Second, we show that a large share of the compliers would have attended another preschool in the absence of the program. Interestingly, the presence of alternative preschools is frequent in the preschool literature: all of the existing literature studying preschool impact report similar substitution patterns. In this paper, we argue that failure to identify the sub-LATE parameter and, in particular, the effect on children who would have stayed at home ($LATE_{hc}$ here) is a shortcoming of the current literature on preschool in low-income countries. We show that, with a set of very plausible assumptions, we can derive bounds for the $LATE_{hc}$ and show that, after about 9 months of preschool, a child's performance increases between 0.13 SD and 0.45 SD. With additional baseline predictors, the bounds can be significantly narrowed to 0.14 and 0.39 SD. Using additional instrumental variables and under heavier assumptions (constant treatment effects), we estimate the $LATE_{hc}$ is positive and significant at around 0.16 SD (between 0.16 SD and 0.32 SD depending on the instruments used). Finally, our predicted LATE approach estimate the $LATE_{hc}$ to be at 0.27

SD on the cognitive development index. This results correspond to an effect of sizable magnitude as it represents about two-third of the cognitive gap measured at baseline between relatively wealthier (top quintile) and poorer children (bottom quintile). In both approaches, the effect of switching from APS to CPS (0.088 for conditional LATE and 0.055 for predicted LATE) is small and our detection capacities, indicating that, while CPS extended the preschool offer in Cambodia, did not improved the quality by much.

Our results directly relate to the existing literature on preschool impact. They are in line with the most positive results reported in Mozambique (Martinez et al., 2017) and are in sharp contrast with the more disappointing results found in comparable studies in Cambodia, The Gambia , and Indonesia (Bouguen et al., 2018, 2013; Blimpo and Pugatch, 2017; Brinkman et al., 2017). While implementation issues and failure to account for substitution patterns may explain some of these previous results, other studies, less concerned by the substitution issue, also raise doubts over the effectiveness of early childcare development programs (Ozler et al., 2018; Andrew et al., 2018). Our results show that a properly implemented preschool provision, designed and conducted entirely by the Cambodian government, impacts the learning capacities of children. The effect of such a policy is particularly large on children who would have otherwise stayed at home. It means that a similar policy implemented in a context where no alternative childcare provision exists would prove to be a very effective education policy.

This article also relates to the literature on close substitute programs and on the identification of sub-LATEs (Kline and Walters, 2016; Heckman et al., 2000; Hull, 2018; Kirkeboen et al., 2016; Feller et al., 2016). We contribute to that literature by showing that the effect on children who would have stayed at home can be bounded. Our bounding strategy can be implemented in many contexts, is reliable, and is based on very plausible assumptions. Our novelty approach, predicted LATE, proved effective to estimate the sub-LATE in our context. With larger datasets and improved machine learning tools, this technique may find itself useful in many context. Yet extracting bounds and implementing alternative identification strategies depend on one important condition: the experiment must be powered to detect effects on children who would have enrolled in any program (here, any preschool take-up). In other words, the program's take-up (here, the CPS take-up), which is typically used in power calculations (Duflo et al., 2008), would generally not provide enough detection power in presence of close substitutes. Including the substitution patterns in the design and power calculation – through pilot studies or a careful analysis of the available substitution offers – is critical for precisely isolating the sub-LATE parameters.

References

Alkire, S. and M. E. Santos (2010). Acute multidimensional poverty: A new index for developing countries. Technical report, Human Development Report Office (HDRO), United Nations Development Programme (UNDP).

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, E. Fitzsimons, S. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in colombia. *PLoS medicine 15*(4), e1002556.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

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.

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.

Berkes, J., A. Bouguen, and D. Filmer (2017, 10). *Increasing Early Childhood Care and Development Through Community Preschools in Cambodia: Evaluating the Impacts.* `https://doi.org/10.1257/rct.1045-3.0`: AEA RCT Registry.

Berkes, J., A. Bouguen, D. Filmer, and T. Fukao (2019). Improving preschool provision and encouraging demand : Heterogeneous impacts of a large-scale program. Impact evaluation series, Washington, D.C. : World Bank Group.

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.

Blimpo, M. P. and T. Pugatch (2017). Scaling up children'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.

Bouguen, A., M. Gurgand, and J. Grenet (2017). Does class size influence student achievement? Technical Report 28, PSE.

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.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). Early childhood education: Young adult outcomes from the abecedarian project. *Applied Developmental Science 6*(1), 42–57.

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.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Technical report, National Bureau of Economic Research.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *The American Economic Review 97*(2), 31.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Currie, J. (2001). Early childhood education programs. *The Journal of Economic Perspectives 15*(2), 213–238.

Dean, J. T. and S. Jayachandran (2020). Attending kindergarten improves cognitive development in india, but all kindergartens are not equal.

Duflo, E., R. Glennerster, and M. Kremer (2008, January). *Using Randomization in Development Economics Research: A Toolkit*, Volume 4 of *Handbook of Development Economics*, Chapter 61, pp. 3895–3962. Elsevier.

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.

Fernald, L. C., E. Prado, P. Kariger, A. Raikes, et al. (2017). A toolkit for measuring early childhood development in low and middle-income countries. *World Bank Publications*.

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.

Grantham-McGregor, S. M., C. A. Powell, S. P. Walker, and J. H. Himes (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the jamaican study. *The Lancet 338*(8758), 1–5.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics 14*(3), 262–280.

Heckman, J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature 48*(2), 356–98.

Heckman, J., N. Hohmann, J. Smith, and M. Khoo (2000). Substitution and dropout bias in social experiments: A study of an influential social experiment*. *The Quarterly Journal of Economics 115*(2), 651–694.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

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*.

Jung, H. and A. Hasan (2016). The impact of early childhood education on early achievement gaps in indonesia. *Journal of Development Effectiveness 8*(2), 216–233.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection*. *The Quarterly Journal of Economics 131*(3), 1057–1111.

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.

Martinez, S., S. Naudeau, and V. Pereira (2017). The promise of preschool in africa: A randomized impact evaluation of early childhood development in rural mozambique. *Washington, DC: The World Bank*.

Meager, R. (2018, June). Understanding the Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of 7 Randomised Experiments. *American Economic Journal: Applied Economics (forthcoming)* (1506.06669).

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.

Ozler, B., L. C. Fernald, P. Kariger, C. McConnell, M. Neuman, and E. Fraga (2018). Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial. *Journal of Development Economics 133*, 448 – 467.

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.

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221.

Schady, N., J. Behrman, M. C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis (2015). Wealth gradients in early childhood cognitive development in five latin american countries. *Journal of Human Resources 50*(2), 446–463.

UNESCO, Brookings Institution, W. B. (2017). Overview melqo: Measuring early learning quality outcomes.

Walker, S. P., S. M. Chang, M. Vera-Hernández, and S. Grantham-McGregor (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics 127*(5), 849–857.

Figures

Figure 1: State Preschool (SPS)

Note: State preschools are generally attached to a primary school and classes are given by a formal preschool teacher. State preschools have usually more experienced teachers and state preschool teachers are better paid than community teachers. Classes in State preschool last 3 hours, 5 days a week against 2 hours in CPSs or IPSs.

Figure 2: Informal Preschool (IPS)

Note: pictures of an informal preschool (IPS) classroom, usually given at the community teacher's house (here under her house).

Figure 3: Community Preschool (CPS)

Note: pictures of a community preschool (CPS). CPSs were built under GPE II. CPSs are all standard: they are usually much better equipped than informal preschools. Newly recruited teachers receive better training and usually higher wages. Class lasts for 2 hours each day.

Figure 4: Enrollment before and After School Construction

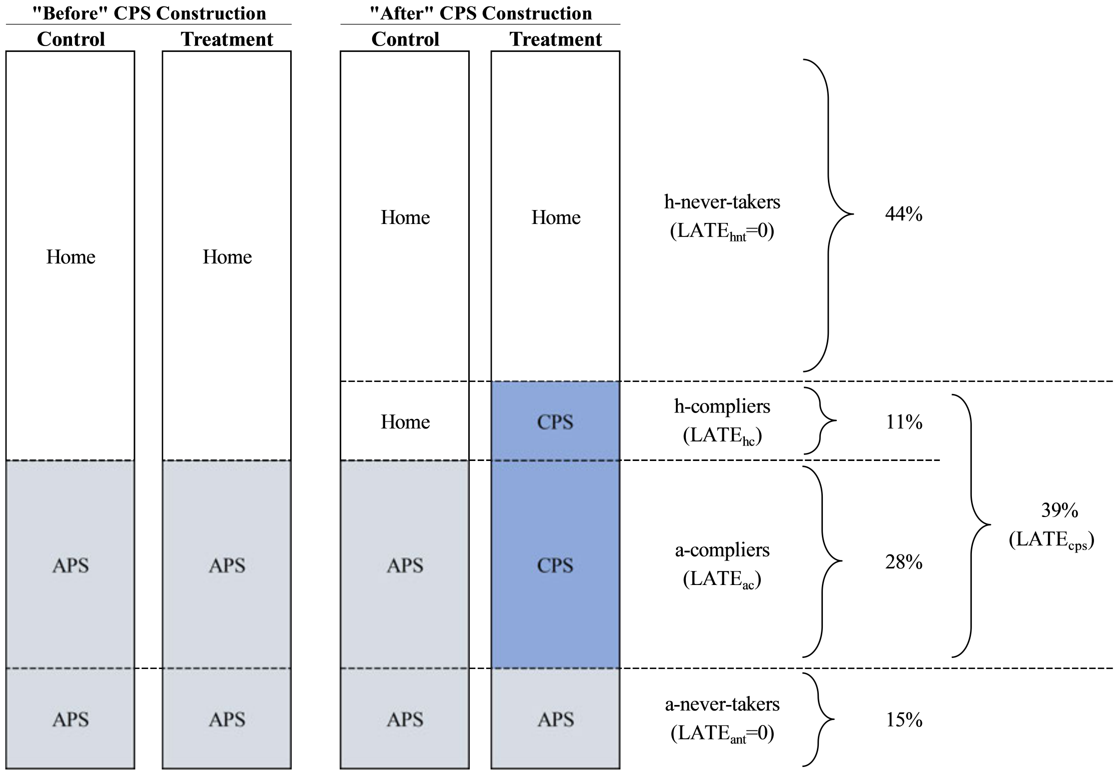

Figure 4 shows care arrangements ($D \in \{c, a, h\}$) of children in treatment and control groups. The left panel shows the counterfactual scenario in the absence of the program. The right panel shows the observed scenario at follow-up under implementation of the program. Randomization implies that the control group at follow-up is equivalent to the treatment group at follow-up in the absence of the program.

Figure 5: Predicted LATE approach

Figure 5 gives a visual representation of the predicted LATE approach to identify $LATE_{hc}$ and $LATE_{ac}$. After having computed the predicted probability of enrolling in APS using LASSO selected baseline variables, the approach consists in identifying the S_{ac} % (respectively $1 - S_{ac}$ %) CPS enrollees with the highest (resp. lowest) predicted probability as a-compliers (resp. h-complier). $LATE_{ac}$ (resp. $LATE_{hc}$) is then estimated using the sub-sample composed of APS enrollees (resp. children who stayed at home) and predicted a-compliers (h-compliers) via 2SLS.

Figure 6: $\hat{D}_a$ distribution and compliers type - CPS children at follow-up

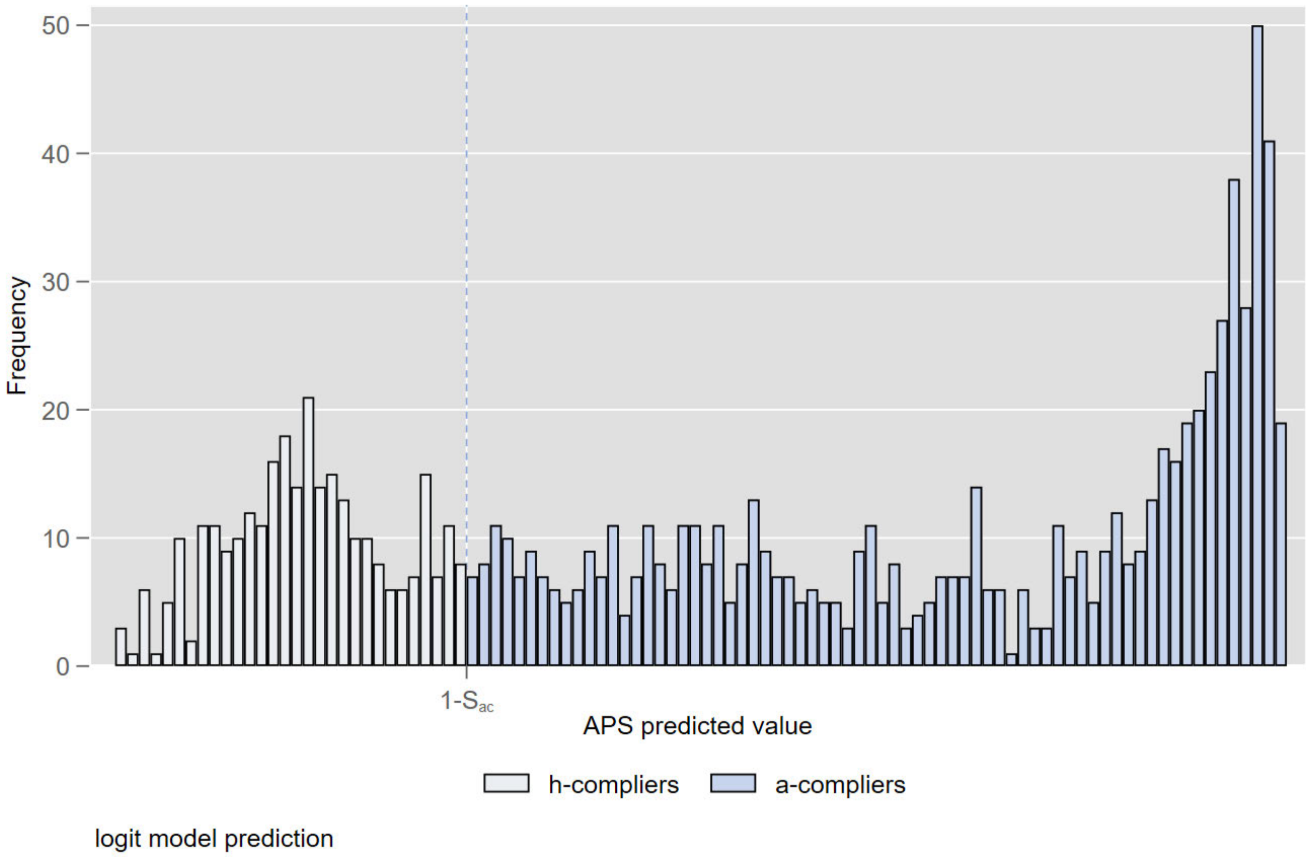

Figure 6 gives the distribution of $\hat{D}_a$, the APS enrollment's predicted value computed from equation 12 and the identification of the a- and h-compliers using the approach described in 3.6.

Tables

Table 1: Timetable

Period	**Activity**	**CPS construction**
03/2016	Begin CPS construction	0% completed
05/2016 - 07/2016	Baseline data collection	17% completed
10/2016	School entry	?
12/2016	Monitoring survey (by phone)	72% completed
04/2017 - 06/2017	Follow-up data collection	86% completed

Note: Percentages refer to share of villages in the treatment group for which construction of a new CPS was completed at the day of data collection.

Table 2: Study Sample

	Total	Attrition rate	Treatment	Control	Differential attrition
Baseline May-July 2016					
Villages	178		120	58	
Households	4115		2839	1276	
Household members	22240		15347	6893	
children from 2 -4	4393		3058	1335	
Tested children	4316		3008	1308	
Midline April-June 2017					
Villages	178		120	58	
Households	3757		2578	1179	
Household members	20485		14080	6405	
children from 3-5	4018		2762	1256	
Tested children	3963		2721	1242	
Baseline & midline					
Villages	178		120	58	0.0%
Households	3718	13.9%	2572	1146	2.1%
Households members	20283	8.8%	14045	6238	-1.0%
children from 3-5	3973	9.6%	2751	1222	1.6%
Tested children	3857	10.6%	2671	1186	1.9%

The table provides the study universe in term of villages, households, eligible children, and tested children at baseline and at follow-up (1 year after baseline). The attrition and differential attrition columns give the respective overall attrition rate and the differential between treatment and control attrition.

Table 3: School construction and Children Enrollment at follow-up

	Obs.	C	T-C
School construction			
Any preschool	178	0.81	0.123**
			(0.057)
Community preschool (CPS)	178	0	0.858***
			(0.032)
Alternative preschool (APS)	178	0.81	-0.552***
			(0.066)
··· Informal (IPS)	178	0.655	-0.564***
			(0.068)
··· State (SPS)	178	0.241	-0.058
			(0.067)
Children Enrollment			
Any school	4011	0.435	0.106***
			(0.036)
Any school exposure (m)	4006	3.672	1.01**
			(0.401)
Community preschool (CPS)	4011	0	0.389***
			(0.025)
Alternative preschool (APS)	4011	0.435	-0.283***
			(0.034)
··· Informal preschool (IPS)	4011	0.284	-0.247***
			(0.034)
··· State preschool (SPS)	4011	0.112	-0.041*
			(0.024)
··· Primary school	4011	0.04	0.004
			(0.009)

Table 3 presents village-level and children-level regressions of the outcome variable in rows against the treatment variable. Enrollment takes value 1 when the child is currently enrolled in school. Exposure gives the total number of month the child ever spent in (any) school. Estimates correct for heteroskedasticity and, for children-level data are cluster at the village level. Column *T-C* gives the result of the regression without any control, Column *C* the average in the control, and Column *Obs.* the number of observations.
* 10%, ** 5%, *** 1 % significance level

Table 4: ITT and $LATE_{hc}$ Bounds

		ITT		**$LATE_{hc}$ Bounds**		**$LATE_{hc}$ Narrow bounds**	
	Obs.	(1)	(2)	Lower	Upper	Lower	Upper
Any School Exposure (m)	4006	1.068***	1.069***	2.710***	9.537***	2.444***	8.287***
		(0.364)	(0.352)	(0.815)	(1.672)	(0.853)	(1.329)
Executive functions	3959	0.055	0.050*	0.126*	0.439*	0.134**	0.371
		(0.038)	(0.026)	(0.065)	(0.224)	(0.065)	(0.231)
Language	3959	0.055	0.046	0.117	0.408	0.134*	0.348
		(0.040)	(0.030)	(0.075)	(0.267)	(0.078)	(0.237)
Numeracy	3959	0.048	0.049*	0.127*	0.441	0.136*	0.204
		(0.038)	(0.029)	(0.072)	(0.271)	(0.078)	(0.229)
Fine motor	3959	0.077*	0.061**	0.152**	0.527*	0.167**	0.38
		(0.042)	(0.030)	(0.075)	(0.271)	(0.077)	(0.251)
Socio-emotional	3959	0.039	0.050	0.127	0.442	0.132	0.395
		(0.037)	(0.037)	(0.092)	(0.342)	(0.095)	(0.338)
Cognitive development index	3959	0.055*	0.051**	0.130**	0.451**	0.14***	0.339*
		(0.030)	(0.020)	(0.051)	(0.195)	(0.052)	(0.176)

Table 4 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 variable ($\boldsymbol{W}$). Narrow bounds are estimated using $\boldsymbol{W}$ and $\boldsymbol{B}$ as control variables and instrument the endogeneous variable (CPS enrollment or any preschool enrollment) by Z, $\boldsymbol{B}$ and $\boldsymbol{B}$*Z. $\boldsymbol{B}$ includes a dummy for each quintile of $\hat{S}_{ac}(\boldsymbol{X})$ the predicted share of a-compliers. Standard errors are robust to heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Table 5: Bounds and Conditional LATE

	Bounds		**Prov. FE & village char.**		**Prov. FE, vill. & hh char.**	
	Lower	Upper	$LATE_{hc}$	$LATE_{ac}$	$LATE_{hc}$	$LATE_{ac}$
School exposure (months)	2.710***	9.537***	8.558***	0.446	8.666***	0.417
	(0.815)	(1.672)	(0.802)	(0.534)	(0.742)	(0.522)
Overid. test p-value			0.686		0.928	
Executive functions	0.126*	0.439*	0.11	0.08	0.226**	0.035
	(0.065)	(0.224)	(0.126)	(0.075)	(0.115)	(0.073)
Overid. test p-value			0.128		0.066	
Language	0.117	0.408	0.129	0.071	0.201*	0.051
	(0.075)	(0.267)	(0.122)	(0.094)	(0.111)	(0.087)
Overid. test p-value			0.835		0.818	
Numeracy	0.127*	0.441	0.105	0.075	0.191*	0.058
	(0.072)	(0.271)	(0.118)	(0.091)	(0.113)	(0.089)
Overid. test p-value			0.187		0.114	
Fine motor	0.152**	0.527*	0.055	0.172**	0.054	0.17**
	(0.075)	(0.271)	(0.163)	(0.083)	(0.152)	(0.08)
Overid. test p-value			0.844		0.965	
Socio-emotional	0.127	0.442	-0.009	0.172	0.117	0.127
	(0.092)	(0.342)	(0.17)	(0.118)	(0.161)	(0.121)
Overid. test p-value			0.478		0.17	
Cognitive Development	0.130**	0.451**	0.078	0.114*	0.158*	0.088
	(0.051)	(0.195)	(0.101)	(0.061)	(0.089)	(0.057)
Overid. test p-value			0.327		0.155	
Observation	3959	3959	3959	3959	3959	3959
Underid p-value			0.000	0.000	0.000	0.000

Table 5 first provides the bounds from Table 4 in the first two columns. Then, we provide the estimates of $LATE_{hc}$ and $LATE_{ac}$ based on the approach described in Section 3.5: we instrument the endogenous variables by a set of baseline variables, $\boldsymbol{X}$, Z and their interactions. $\boldsymbol{X}$ is composed of the province fixed effect, and village characteristics in column 3 and 4 to which we add households characteristics in column 5 and 6. For each conditional LATE estimation, we provide the p-value of the over-identification test (Hansen et al., 1996) and p-value of the underidentification test ((Sanderson and Windmeijer, 2016)) of the first stage regressions in the last row. Standard errors are robust to heteroskedasticity and clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Table 6: Predicted $LATE_{hc}$ using LASSO

	Bounds		**Conditional LATE**		**Predicted LATE**	
	Lower	Upper	$LATE_{hc}$	$LATE_{ac}$	$LATE_{hc}$	$LATE_{ac}$
School exposure (months)	2.710***	9.537***	8.666***	0.417	5.068***	1.268*
	(0.815)	(1.672)	(0.742)	(0.522)	(0.217)	(0.654)
Executive functions	0.126*	0.439*	0.226**	0.035	0.287*	0.084
	(0.065)	(0.224)	(0.115)	(0.073)	(0.148)	(0.059)
Language	0.117	0.408	0.201*	0.051	0.248*	0.054
	(0.075)	(0.267)	(0.111)	(0.087)	(0.134)	(0.077)
Numeracy	0.127*	0.441	0.191*	0.058	0.085	0.094
	(0.072)	(0.271)	(0.113)	(0.089)	(0.147)	(0.075)
Fine motor	0.152**	0.527*	0.054	0.17**	0.283*	0.054
	(0.075)	(0.271)	(0.152)	(0.08)	(0.147)	(0.073)
Socio-emotional	0.127	0.442	0.117	0.127	0.464*	-0.014
	(0.092)	(0.342)	(0.161)	(0.121)	(0.238)	(0.073)
Cognitive Development	0.13**	0.451**	0.158*	0.088	0.273***	0.055
	(0.051)	(0.195)	(0.089)	(0.057)	(0.098)	(0.050)
Observation	3,959	3,959	3,959	3,959	2,264	1,747

The table compares the results from the bounding strategy, from the conditional LATE and last from the predicted LATEs. The predicted $LATE_{hc}$ consists in first identifying the group of h-compliers within the group of compliers using machine learning. Then, the $LATE^{p}_{hc}$ is estimated via 2SLS in the subgroup of the h-compliers and h-never-takers. Similarly, the predicted $LATE_{ac}$ consists in first identifying the a-compliers within the compliers and then estimating the $LATE^{p}_{ac}$ via 2SLS within the group of a-compliers and a-never-takers. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Appendix A- Complementary analysis

Table A1: Treatment-Control Difference at Baseline – Household and Caregiver Data

	Baseline Sample			Baseline & Follow-up Sample		
	Obs.	C	T-C	Obs.	C	T-C
Household characteristics						
Household size	4115	5.402	0.004	3718	5.443	0.017
			(0.097)			(0.1)
Multidimensional poverty	4115	0.412	0.002	3718	0.393	0.004
			(0.032)			(0.031)
House is rented	3560	1.08	-0.002	3202	1.076	-0.007
			(0.015)			(0.015)
Income > $100	4115	0.452	-0.018	3718	0.476	-0.036
			(0.039)			(0.039)
No one completed prim. school	4115	0.221	0.014	3718	0.224	0.002
			(0.025)			(0.026)
Farming activity	4074	0.825	0.005	3716	0.838	0
			(0.029)			(0.029)
Caregivers characteristics						
Female	4391	0.89	0.019	3916	0.89	0.019
			(0.013)			(0.013)
Age	4391	40.777	-0.227	3916	40.669	-0.104
			(1.014)			(1.017)
# of years of education	4330	4.16	-0.216	3868	4.165	-0.173
			(0.239)			(0.25)
Biological parent	4333	0.596	-0.008	3866	0.602	-0.011
			(0.034)			(0.035)
Malnourished	4371	0.141	0.011	3897	0.141	0.009
			(0.015)			(0.016)
Ravenscore (cognitive test)	4344	0.05	-0.107	3872	0.048	-0.095
			(0.067)			(0.07)
Cognitive parenting score	4379	-0.006	0.017	3906	0.023	-0.01
			(0.056)			(0.058)
Negative parenting score	4380	0.002	0.059	3907	0.009	0.043
			(0.062)			(0.063)
Socio-emotional parenting score	4379	-0.008	-0.026	3906	0.022	-0.057
			(0.048)			(0.049)

Each line represents a regression of an outcome variable on treatment group indicators. The first panel looks at the data collected at baseline, while the second at the data collected at baseline among individuals present at follow-up. Estimates correct for heteroskedasticity and intra-village correlations.
* 10%, ** 5%, *** 1 % significance level

Table A2: Treatment-Control Difference at Baseline – Children Data

	Baseline Sample			Baseline & Midine Sample		
	Obs.	C	T-C	Obs.	C	T-C
Sample children Characteristics						
Age	4393	3.476	0.005	3918	3.485	0.002
			(0.03)			(0.032)
Female	4393	0.506	-0.022	3918	0.506	-0.02
			(0.017)			(0.018)
Child ill in the last month	4380	0.778	0.023	3907	0.782	0.019
			(0.018)			(0.018)
Complete vaccination	4381	0.548	-0.03	3908	0.554	-0.027
			(0.037)			(0.037)
Underweight	4313	0.302	0.012	3852	0.306	-0.001
			(0.02)			(0.02)
Stunting	4299	0.341	0.026	3841	0.335	0.022
			(0.019)			(0.02)
Sample children Score						
Emerging numeracy	4316	0	-0.065	3857	0.009	-0.066
			(0.046)			(0.045)
Language	4316	0	-0.05	3857	0	-0.034
			(0.051)			(0.049)
Executive function	4316	0	-0.004	3857	0.001	0.013
			(0.049)			(0.05)
Fine motor	4316	0	0.027	3857	0.005	0.033
			(0.052)			(0.054)
Gross motor	4316	0	-0.013	3857	-0.008	0.005
			(0.051)			(0.054)
Socioemotional	4303	0	-0.041	3846	0.005	-0.043
			(0.058)			(0.06)
Pre-program Preschool attendance						
Currently attending preschool	4380	0.153	0.066***	3907	0.152	0.072***
			(0.023)			(0.024)
Days in preschools	4379	35.957	10.968*	3906	35.758	12.697**
			(5.631)			(5.732)
Currently attending IPS or CPS	4380	0.123	0.061***	3907	0.122	0.069***
			(0.023)			(0.024)
Currently attending SPS	4380	0.03	0.005	3907	0.03	0.003
			(0.01)			(0.009)
Home based program	4375	0.104	0.031	3901	0.112	0.023
			(0.02)			(0.022)
Home visit	4375	0.017	0.003	3901	0.018	0.002
			(0.006)			(0.007)

Each line represents a regression of an outcome variable on treatment group indicators and province fixed effect (omitted). Estimates correct for heteroskedasticity and intra-village correlations.
* 10%, ** 5%, *** 1% significance level

Table A3: Baseline Comparison of Informal, Community, and State Preschool

	Obs.	IPS	CPS-IPS	SPS-IPS
General Characteristics				
Used only for preschool	267	0.526	0.232***	0.057
			(0.07)	(0.073)
Class-size	267	20.647	1.353	5.895***
			(1.324)	(2.1)
Preschool material				
Tables, 0/1	266	0.15	0.059	0.596***
			(0.061)	(0.061)
Chairs, 0/1	266	0.211	0.031	0.564***
			(0.065)	(0.061)
Books, 0/1	252	0.711	-0.033	0.058
			(0.073)	(0.066)
Pen, 0/1	256	0.539	0.058	0.219***
			(0.077)	(0.069)
Games, 0/1	259	0.577	-0.061	-0.025
			(0.077)	(0.075)
Blackboard, 0/1	263	0.71	0.032	0.133**
			(0.069)	(0.059)
Sum material, 0/6	267	2.827	0.125	1.395***
			(0.257)	(0.285)
Preschool problems				
Poor building, 0/1	267	0.075	-0.059**	0.022
			(0.028)	(0.042)
Low teachers wage, 0/1	267	0.18	-0.084*	-0.014
			(0.051)	(0.055)
Budget constraint, 0/1	267	0.241	-0.144***	-0.032
			(0.053)	(0.061)
Not enough spots, 0/1	267	0.714	-0.214***	-0.367***
			(0.075)	(0.069)
Not enough supplies, 0/1	267	0.737	-0.076	-0.167**
			(0.072)	(0.07)
Poor teacher quality, 0/1	267	0.06	0.02	-0.005
			(0.04)	(0.034)
Class held irregularly, 0/1	267	0.098	-0.065*	-0.042
			(0.034)	(0.038)
Sum problems, 0/10	267	2.526	-0.317**	-0.596***
			(0.15)	(0.16)
Teacher characteristics				
Any training, 0/1	267	0.955	-0.003	0.003
			(0.033)	(0.03)
Days of training	221	78.9	13.7	152.9***
			(22.733)	(37.474)
Is paid, 0/1	267	0.759	0.176***	0.185***
			(0.049)	(0.046)
Wage, USD	250	35.185	3.797	50.856***
			(2.362)	(8.178)

Baseline comparison between the three type of preschool types available in Cambodia (IPS, SPS and CPS), according to the village chief questionnaire. Based on the full sample of 267 schools at baseline.
* 10% significance level ** 5% significance level *** 1% significance level

Table A4: Follow-up Comparison, Informal, Community, and State preschools

	Obs.	IPS	CPS-IPS	SPS-IPS
General preschool characteristics				
Used for preschool only	339	0.627	0.357***	-0.209**
			(0.061)	(0.085)
Open since, days since 1960	279	19138	1532.4***	-2994.4***
			(283.661)	(897.24)
Preschool problems				
Poor building, 0/1	339	0.729	-0.517***	-0.464***
			(0.061)	(0.077)
Too many children, 0/1	339	0.407	-0.175***	-0.195**
			(0.066)	(0.078)
Not enough teacher, 0/1	339	0.237	0	-0.009
			(0.061)	(0.077)
Not enough training, 0/1	339	0.407	0.077	-0.309***
			(0.069)	(0.072)
Not enough tables & chairs, 0/1	339	0.678	-0.405***	-0.443***
			(0.068)	(0.084)
Not enough teaching material, 0/1	339	0.814	-0.101*	-0.172**
			(0.058)	(0.082)
No sanitary facility, 0/1	339	0.593	0.197***	-0.438***
			(0.067)	(0.08)
No clean water, 0/1	339	0.678	0.064	-0.316***
			(0.066)	(0.087)
Class held irregularly, 0/1	339	0.288	0.015	-0.132*
			(0.065)	(0.075)
Other, 0/1	339	0.051	0.021	-0.018
			(0.032)	(0.036)
Sum problems, 0/10	339	4.881	-0.825**	-2.497***
			(0.323)	(0.38)
Teacher characteristics				
Is paid, 0/1	339	0.966	0.006	-0.019
			(0.025)	(0.037)
Paid regularly	339	0.915	-0.009	0.033
			(0.039)	(0.045)
Wage, USD	326	44.5	0.22	132.5***
			(6.474)	(11.313)
# of working days	336	5.103	-0.024	0.219**
			(0.088)	(0.095)
# of teachers at school	338	1.034	-0.018	0.073
			(0.025)	(0.053)

Follow-up comparison between the three types of preschool types available in Cambodia, according to the village chief questionnaire (1 questionnaire per preschool). Based on the full sample of 339 schools.
* 10% significance level ** 5% significance level *** 1% significance level

Appendix B - ITT and $LATE_{cps}$ assumptions

The identification of equation 1 first relies on the assumption A1 that Z is independent of D and Y. Tables A1 and A2 confirm that no imbalances on observable characteristics occur; hence, we consider the randomization as being successful and Z as independent of D and Y (Assumption A1).

Additionally, identification requires absence of spill-over effects across treatment and control group villages (Assumption A2[39]). While a few treatment group villages are in the vicinity of control group villages, we have no reason to believe that the construction of a CPS had any impact on the education provision of children in the control group. For instance, no control group children attended a CPS at baseline or follow-up. Further, CPS teachers are almost always hired from the same village and, thus, their recruitment is not related to the availability of teachers in the control group.

In addition to assumptions A1 and A2, the identification of $LATE_{cps}$ relies on the first-stage assumption (A3) or non-zero average causal effect of Z on D (Angrist et al., 1996). We verify assumption A3 using the following first-stage regression:

$$\mathbb{1}_{\{D_i=c\}} = \alpha_0^c + \alpha_{FS}^c Z_v + \boldsymbol{W_i \alpha^c} + \mu_v^c + \epsilon_{iv}^c \tag{14}$$

where $\mathbb{1}_{\{D_i=c\}}$ takes 1 when the child is enrolled in CPS and 0 otherwise and A3 is respected when $\alpha_{FS} > 0$. Table 3 show that this is valid assumption in our case.

In the presence of close substitutes, we reformulate the exclusion restriction in the following way:

Assumption 4. - *Exclusion Restriction (A4)*

$$Y_i(c,1) = Y_i(c,0) \tag{A4.1}$$
$$Y_i(a,1) = Y_i(a,0) \tag{A4.2}$$
$$Y_i(h,1) = Y_i(h,0) \tag{A4.3}$$

As we show in the result sections, Case A4.1 can be ruled out by construction in our context since no control children attend a CPS. Assumption A4.2 and A4.3, are subject to violations if the construction of a CPS affect the performance of the never-takers (a or h). CPS construction may reduce APS class-size, change APS peer composition, or make more salient to parents the importance of early educative investment, affecting the performance of both a and h never-takers. We have reasons to believe this is unlikely to be the case in our context.

First, in term of class-size, we have reasons to believe that this problem is unlikely. Indeed, as seen in Table 3, the construction of a CPS generally means that the IPS shuts down: only 7 treatment villages kept their IPS when a CPS was constructed (6% of the treatment villages). This means that in 94% of the cases, the CPS did not have the indirect effect of reducing IPS class size: thus, IPSs are unlikely to have indirectly benefited from class size reduction.[40] Yet, since SPSs were not shut down when a CPS was constructed, SPSs are more likely to have been indirectly affected by the CPS construction. We look at this possibility in the last row of Table 3. Since we did not collect class sizes in SPSs, as a proxy, we use the

[39]We do not use the traditional SUTVA assumption because, for the identification of ITT, the absence of spill-over (or general equilibrium effect) across treatment branches is sufficient. SUTVA, as described by Angrist et al. (1996), has larger implications that will be covered in A4.

[40]Note that we cannot test the IPS class-size because IPSs were closed down when a CPS opened . As a result IPSs are not comparable in treatment and control.

average number of sampled children enrolled in SPS class per village. Since we did not sample all children in the village (but an average of 26 children), we divide this number by our average sample weight (here estimated at 53%). The last row of Table 3 indicates that, on average, the number of sampled children enrolled in SPS in treatment group is 1 unit lower than in the control group. The point estimate is not significant but would correspond, if taken at face value, to a class size reduction of about 1.83 children (0.97/0.53). Since SPS enrollment concerns about 8% of the sampled children (0.083), and since the impact of class size is reported in the literature to be maximum -3 pp per additional students (Bouguen et al., 2017), the indirect effect on class size reduction is estimated to be $1.83 * 0.083 * 0.03 = 0.4$ pp maximum. This would correspond to about 8% of the overall treatment effect (5.1% of a SD). Hence, reduction of class size in SPS is unlikely to significantly modify the magnitude of the treatment effect.

Second, peer composition may violate the exclusion restriction if, for instance, CPSs attract specific children, leaving SPSs or the remaining IPSs with more homogeneous or better/worst peers. As mentioned in the body of the text, since APSs are composed of better quality schools, we would expect the peer composition to have improved in APSs and, as a result, the a-never-takers to benefit from more favorable conditions. We test this possibility in Table B1 for SPS children, where we look at baseline balancing for the SPS children at follow-up. We do not find any significant difference between treatment and control in terms of baseline characteristics. This suggests that CPS constructions did not modify a-never-takers' composition.[41]

Lastly, the exclusion restriction may be violated if CPS construction modified the involvement of never-taker parents (h and/or a). While we report elsewhere (Berkes et al., 2019) that the program positively impacts parents' perceptions and self-reported parenting practices, such effect does not constitute an A4 violation as long as it only affects the a- and h-compliers.[42] A4 would be violated, however, if the parenting effect expands to h-and a-never-takers.

We look at this possibility in Table B2, where we estimate the ITT effect on SPS never-takers children. Recall that enrollment in SPS is not affected by the treatment and therefore both population are comparable. Table B2 shows that parents do not report different perceived returns to education, that children do not perform better, but that parents do declare being more involved, on average, in their children's education in the treatment. Since parenting scores are self-reported, this could simply be a reporting bias: with the construction of a preschool in the village, all parents are more inclined to report positive parenting behaviors, while their actual parenting involvement might not have been significantly modified. Yet, as suggested in the body of the text, it could also be that all parents changed their behavior toward early education because of the construction. If that were the case, it would be a violation of A5. We should not overestimate the magnitude of the problem, however. First, the parenting results are driven by socio-emotional parenting – the dimension least correlated with cognitive performance, according to (Berkes et al., 2019), while cognitive parenting, the parenting measure with the highest predictive power, is of lower magnitude and non-significant (+11 pp). Second, even if we take the cognitive parenting at face value, the potential bias remains minimal. According to Berkes et al. (2019), children's performance increases by a maximum of 10 pp for

[41]Again, the same cannot be done for IPSs, as CPS construction forced many IPSs to close (see Table 3) and, therefore, the IPS treatment sample is a selected one.

[42]Remember that our experiment measures the overall effect of a preschool construction, including indirect effects on parental perception and involvement.

each standard deviation increase of the cognitive parenting index. Hence, an 11 pp effect would translate into a $0.1 * 11 = 0.11$ *pp* effect on children's performance. Since this effect applies to only 8.3% of all children, the potential effect of the violation of the assumptions is infinitesimal ($0.083 * 0.011 = 0.09$pp compared with an overall effect of 5.1 % of a SD). Given the low magnitude the potential bias, we really do not believe it is a cause of concern for our experiment.

In all, the fact that we do not find any positive impacts for children in Table B2 is evidence that Assumption A4 is valid on the whole. Indeed, class-size, peer composition, and parental involvement are all forces that would bias upward the impact in Table B2 in case of violation of A4. With an overall ITT effect on SPS children estimated at -0.02 SD, we are confident that our experiment is not affected by a violation of the exclusion restriction.

Table B1: Baseline Description of Children Enrolled in SPS at Follow-up

	Obs.	C	T-C
Household Characteristics			
Household size	336	5.279	0.323
			(0.211)
Multidimensional poverty index	336	0.314	0.022
			(0.082)
Farmer	330	0.866	-0.121*
			(0.065)
No one > 5 years of education	336	0.171	0.022
			(0.052)
Caregivers characteristics			
Raven score (cognitive test)	328	0.139	0.025
			(0.147)
Cognitive parenting score	329	0.26	-0.197
			(0.154)
Negative parenting score	329	0.004	0.086
			(0.134)
Socio-emotional parenting score	329	0.194	-0.077
			(0.11)
Children characteristics			
Early Numeracy	322	0.456	-0.177
			(0.116)
Language	322	0.537	-0.135
			(0.14)
Executive functions	322	0.504	0.003
			(0.168)
Fine motor	322	0.343	0.065
			(0.138)
Socio-emotional	322	0.346	0.129
			(0.131)
Gross motor	321	0.305	-0.101
			(0.139)

Table B1 presents children level regressions of the outcome variable in line against the treatment variable. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1 significance level

While A4.1, 2 and 3 are necessary conditions for the $LATE_{cps}$ to be identified, they are not sufficient without an extended monotonicity assumption that takes close substitutes into account.

Table B2: ITT estimate on Children Enrolled in SPS at Follow-up

	Obs.	C	T-C
Cognitive Development (CD)			
Executive functions	332	0.544	0.11
			(0.071)
Language	332	0.664	-0.037
			(0.097)
Numeracy	332	0.677	-0.065
			(0.093)
Fine motor	332	0.739	-0.144
			(0.093)
Socio-emotional	332	0.344	0.027
			(0.094)
CD index	332	0.594	-0.022
			(0.051)
Parenting Score (PS)			
Negative parenting	336	-0.04	-0.109
			(0.107)
Socioemotional parenting	336	-0.129	0.216*
			(0.122)
Cognitive parenting	336	0.154	0.114
			(0.147)
PS Index	336	0.022	0.146*
			(0.088)
Parental Perception			
Optimal preschool age	336	3.821	-0.091
			(0.084)
Optimal Primary school age	336	5.829	-0.06
			(0.065)
Perceived Income no school	336	106.183	-9.848
			(7.538)
Perceived Income Prim. School	336	144.429	6.804
			(11.1)
Perceived Income Sec. School	336	227.777	12.6
			(21.687)

Table B2 presents children level regressions of the outcome variable in line against the treatment variable. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1 significance level

Assumption 5. - *Extended Monotonicity Assumption (A5)*
No child belongs to one of the following strata:

$$\textit{ch-defiers: } D_i(0) = c,\ D_i(1) = h \tag{A5.1}$$

$$\textit{ca-defiers: } D_i(0) = c,\ D_i(1) = a \tag{A5.2}$$

$$\textit{ah-defiers: } D_i(0) = a,\ D_i(1) = h \tag{A5.3}$$

$$\textit{ha-defiers: } D_i(0) = h,\ D_i(1) = a \tag{A5.4}$$

Cases A5.1 and A5.2 are both analogous to defiers in the traditional LATE framework.(Angrist et al., 1996) Since enrollment into CPSs is zero in our control group, we can rule out these two cases and consider valid the traditional monotonicity assumption.

Yet, cases A5.3 and A5.4 are theoretically possible. A5.3 (respectively A5.4) corresponds to the situation, where the CPS construction would either decrease (resp. increase) APS attendance. While the existence of ah-defiers is very unlikely, as the construction of CPSs is unlikely to reduce the overall demand for preschool, ha-defiers deserve more attention.[43] If, for instance, CPS construction positively modifies the perception of preschool in general and entices some parents to enroll their children in APS instead of CPS (because of shorter distance or because the CPS have no additional capacities), then A5 would be violated. Since CPS cater to a maximum of 25 children, excess demand for CPS may result in higher APS attendance. Relatedly, if APSs are already at capacity when the CPS opens, children switching from APS to CPS would make room for ha-defiers. This situation would again violate A5.[44]

However, in our context, the presence of ha-defiers is unlikely. First, when asked about the reasons why their children are not enrolled at preschool, few parents stated that it was because the preschool was already full. While some parents stated that enrollment was turned down, based on observed class sizes and qualitative interviews with teachers, we interpret this as lack of self-sufficiency and emotional maturity to go to preschool rather than capacity constraints.[45] Second – and perhaps more importantly – in the vast majority of cases, the construction of a CPS caused the IPS to shut down: only 7 IPSs remained open after the 103 CPSs were constructed, for a total of only sixty-six children enrolled in an IPS following the construction of a CPS. In the vast majority of cases, children staying at home simply could not have enrolled in an IPS because the IPS no longer existed. This applies to IPSs, not SPSs. SPSs remained open regardless of the treatment status of the village. Therefore, we could be in the presence of ha-defiers if, for instance, the children switching from SPS to CPS could be replaced by children who would have otherwise stayed at home. Again, we do not believe this is likely: SPSs provide a much better education environment – in terms of teacher quality, equipment, and even peers than CPS; see Tables A3 and A4. Thus, parents lacked a reason to remove their children from an SPS and to enroll them into a CPS and, as a result, CPS did not create room in SPSs for ha-defiers

[43]We treat the *ha-defiers* under our extended monotonicity assumption, while Feller et al. (2016) treats it as a sub-assumption called "irrelevant alternatives".

[44]Kline and Walters (2016) discuss this issue for Head Start, where assignment to program preschools could make rationed slots in non-program preschools available to non-treated children. We refer to the same issue as a violation of the extended monotonicity assumption.

[45]see (Berkes et al., 2019) for more details about reasons for not attending