Heterogeneous Preschool Impact and Close Substitutes: Evidence from a Preschool Construction Program in Cambodia*

Jan Berkes †Adrien Bouguen ‡

Adrien Bouguen's 2019 Job Market Paper

November 5, 2018

Abstract

We study the impact of preschools and the issue of close substitutes in a Cambodian context where newly built formalized preschools are competing with existing alternative early childcare arrangements. In addition to estimating the reduced-form impact of a vast preschool construction program using a random assignment, we implement several empirical techniques to isolate the impact on children who would have stayed at home if they had not been enrolled in the newly built preschools. We argue that this parameter is both critical for the preschool literature and, because it does not depend on the quality of alternative preschool, is often more externally valid than standard treatment parameters. Our results show that after one year of experiment, the average Intention-To-Treat impact on cognitive and socioemotional development measures is significant but small in magnitude (0.05 SD). Our analysis, however, suggests that the impact on the children who would have stayed at home will likely be high and significant, between 0.13 SD and 0.45 SD. In a context where infrastructures are improving in low-income countries, our analysis suggests that accounting for close substitutes is crucial to produce more external valid statements on programs' performance and make appropriate policy recommendations.

JEL classification: I24, I25, J24

Keywords: Early Childcare Development (ECD), Supply of Education, Preschool, Close Substitutes, subLATEs.

*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, and Marc Gurgand.

†DIW Berlin

‡Agriculture and Resources Economics Department, University of California, Berkeley

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 close substitutes can affect the interpretation of the standard treatment effect parameters – intention-to-treat (ITT) and local average treatment effect (LATE) – and cloud 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). Even in the presence of close substitutes, standard treatment effect parameters remain internally valid and relevant estimates of the overall policy impact. Yet, clarifying how the effect of a policy depends on close substitutes is critical to produce a more external valid statement on a program's performance and make appropriate policy recommendations.

We study the issue of close substitution in a Cambodian context, where newly built formalized preschools (later referred to as community preschools or CPS) are competing with existing alternative childcare arrangements. In this program – conducted with the World Bank and the Cambodian Ministry of Education, Youth and Sports (MoEYS) – the construction of CPS was randomly assigned among villages with alternative forms of preschool (later referred to as alternative preschool or APS). The study, therefore, creates two sub-populations of compliers: children who would have stayed in home care in the absence of the program (home compliers) and children who would have attended alternative preschools (alternative compliers). Consequently, the ITT and traditional LATE effects presented in this paper measure the effectiveness of the new community preschools in comparison to a mix of home care and alternative preschools.

The presence of close substitute programs is not a unique characteristic of our trial. To our knowledge, every large-scale randomized controlled trial conducted to measure preschool effects in a low-income country was implemented in an environment where alternative care arrangements were present. For instance, in a previous preschool experiment conducted in Cambodia from 2008 to 2010, Bouguen et al. (2018) found that 11% of the control group attended a preschool. Similarly, 8% attended preschool in the control group in Mozambique (Martinez et al., 2017), 16% did so in Gambia (Blimpo and Pugatch, 2017), and 15% in Indonesia (Brinkman et al., 2017). In the US, 40% of the families that lost the 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 a general conclusion on 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 single preschool study.

The contribution of this paper, therefore, is twofold. First, we provide an important contribution to the reduced-form literature on preschool impact by using a large and well-implemented preschool construction program conducted in Cambodia from 2015 to 2018. Second, using new empirical strategies, we isolate the impact of the program on home compliers. We argue that this is a critical parameter in the early childhood development (ECD) literature, and failure to isolate this parameter has thus far contributed to clouding the debate on the effectiveness of ECD in low-income countries.

We start our analysis by looking at the reduced-form estimates. One year after construction of the community preschools, the ITT effect on three- to five-year-old children varies from 0.049 to 0.061 standard deviations (SD) on a large set of child development measures (executive function, language, numeracy, fine-motor, and socio-emotional development). With the exception of socio-emotional development, all ITT effects are statistically significant. While small in magnitude, the overall ITT effect is driven by a 39 percentage point (pp) increase in CPS enrollment and a 11pp increase in overall school enrollment. A larger ITT effect among five-year-olds, from 0.078 SD (executive function) to 0.174 SD (numeracy), is driven by a larger take-up rate of 47 pp. Our results show that a well-implemented, at scale ECD program, conducted by Cambodian teachers, Cambodian trainers, and piloted by the National Ministry of Education in Cambodia, is able to significantly improve cognitive and non-cognitive child outcomes, at least in the short run.

We then document a large degree of program substitution, using detailed information about the alternative forms of preschool available to beneficiaries. In the absence of the construction program, many children would have enrolled in other preschool programs (APS). Hence, the reduced-form effects reflect both the treatment effect of CPS attendance on children who would have stayed at home – home compliers' subLATE or $LATE_{hc}$ – but also the effect from enrolling at CPS instead of APS – alternative compliers' subLATE or $LATE_{ac}$.

To further investigate the underlying causal effects of the preschool program, we show that under plausible assumptions, $LATE_{hc}$ can be bounded between the $LATE_{cps}$ – the LATE of going to CPS – and the $LATE_{ps}$ – the LATE of going to any preschool. With these bounds, the effect on home compliers, who attended the new CPS program for about 10 months, varies between 0.13 SD and 0.45 SD. Finally, we use an empirical technique previously applied elsewhere (Kline and Walters, 2016; Hull, 2018) to estimate the exact size of $LATE_{hc}$ and $LATE_{ac}$. We find that, for cognitive development, the $LATE_{hc}$ is most likely closer to the low bound, while the

$LATE_{ac}$ is positive but generally smaller than the $LATE_{hc}$.

We conclude our analysis by looking at the effects among relatively poorer or wealthier households and find that poorer children did not benefit more from the program than non-poor children. While preschools are often thought to be an efficient way to reduce initial cognitive inequalities between children from wealthier and poorer backgrounds (Elango et al., 2015), preschool attendance has not reduced the cognitive gap in our context and may even increase it. Failure to target the most deprived households likely explains this detrimental effect. Wealthier parents are using preschools as a way to accumulate human capital early, while poorer children are often entirely excluded from preschools.[1]

Our article directly relates to a strand of the applied literature that discusses the interpretation of the treatment parameters in 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, $LATE_{cps}$ is a weighted average of subLATEs on compliers with different counterfactual care arrangements, typically home care or alternative preschool care. Yet, subLATEs parameters, such as $LATE_{hc}$ and $LATE_{ac}$, 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 subLATEs 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 $LATE_{cps}$ are the appropriate parameters. In other contexts, however, failure to isolate subLATEs, and especially failure to isolate the effect on home compliers, is an important limitation for at least three main reasons. First, a large share of the ECD literature is founded on models (Cunha et al., 2010; Heckman, 2010; Cunha and Heckman, 2007; Cunha et al., 2013) that heavily rely on the idea that unfavorable early environments at home should be compensated by formalized ECD programs. This idea prevailed in the development of ECD interventions 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.

Second, many influential empirical papers in the early childcare literature also focused on home compliers. The Jamaica study (Grantham-McGregor et al., 1991) for low-income countries and the Perry Preschool Project (Anderson, 2008) for the US, which constitute the empirical foundation for new ECD interventions, implicitly

[1] In a companion paper (Berkes et al., 2018a), we will investigate strategies to solve this problem by looking at two demand-side interventions aiming at attracting poorest children into preschool.

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, $LATE_{hc}$ does not depend on close substitutes, and its external validity can be assessed using commonly available socioeconomic characteristics, e.g. parental education, poverty, or stunting rate.[3] Any aggregate meta-statistic about the effectiveness of ECD interventions that does not take substitution into account will be 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, from a policy analysis perspective, an understanding of the expected substitution patterns and subLATEs is essential to make appropriate policy recommendations. If the share of home compliers is small, or the $LATE_{ac}$ is null or even negative, large treatment effects on home compliers are entirely consistent with, for instance, low or even non-significant overall reduced-form effects. The subLATE analysis informs the policymakers that the same program, targeted to home compliers, would 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 the children who would benefit most from the program.

The rest of the paper will proceed as follows. First, we describe institutional details, the experimental design, and the studied sample. Second, we will present the empirical framework that will be used to analyze the data. We will focus on the relationship between ITT, LATE, and subLATE parameters. Third, we will 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 a fourth section, we will present our estimation of the $LATE_{hc}$: we will discuss the validity of our bounds and try to give a more precise estimation of the $LATE_{hc}$ under stricter assumptions. We will conclude our analysis by looking at the impact of the preschool construction program on inequality, pointing out the potential detrimental effect of a construction program on cognitive inequality.

[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 analyzed by Ozler et al. (2018), who evaluate the impact of a preschool quality improvement on child performance in Malawi and who implicitly measure a $LATE_{ac}$

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

2 Background, Data, and Design

2.1 Recent ECD Program Development in Cambodia

Despite two decades of robust economic growth, Cambodia remains one of the least developed countries in Southeast Asia, with a GDP per capita estimated at $1160 in 2015 ($3300 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 or Vietnam.[4] To increase the capacities and quality of its education system, the Cambodian government received a first grant from the Global Partnership for Education (GPE 1) of $57 million for the period 2008–2012. The government, in cooperation with the World Bank, used part of the resources to invest in the expansion of the national early education system, which is composed of formal preschools, informal preschools, and parental programs. Bouguen et al. (2013) covered GPE I's impact on the ECD system of education in Cambodia; they stressed that the expansion was accompanied by some implementation problems, including low individual take-up and delays in program implementation. Yet, during that first period, preschool enrollment of five-year-olds increased from 40% in 2009 to 56% in 2012 and 66% in 2016, while enrollment of three- and four-year-olds remained at a low 20% and 37% in 2016, respectively (MoEYS, 2014, 2017).

To improve primary school readiness and be on track with the Sustainable Development Goals,[5] the government of Cambodia, with the support of the World Bank, launched another education expansion program for the period 2014–2018. The plan is financially supported by a second GPE fund (GPE II) of $38 million, still administered by the World Bank, with the objective to strengthen the existing foundation of the education system in Cambodia. Of this amount, about $20 million dollars was allocated to ECD programs. This research project focuses on a part of the expansion that includes the construction of community preschools (CPS).

2.2 Formal Community and Alternative Preschool Programs

Before GPE II, two distinct types of public preschools existed in Cambodia: state preschools (SPS) and community preschools. Since community preschools lacked uniform quality standards, we refer to that type of preschool as informal (community) preschools (IPS). In this article, we consider both IPS and SPS as Alternative

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

[5]The SDG 4.2 states that all children should benefit from at least 1 year of pre-primary education by 2030.

preschools (APS).[6,7] GPE II introduced a new type of community preschool with a uniform 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 $200 to teach for three hours a day, five days a week. As almost all SPS are attached to a public primary school, SPS have access to properly equipped classrooms, teaching, play materials, and sanitary facilities.

In contrast, informal community preschools are typically not attached to a primary school. The local community establishes the IPS and covers 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, and most IPS teachers have to rely on other sources of income. IPS teachers are trained for about 35 days by provincial education departments before they begin with their work. Teachers are required to provide a 2-hour preschool class, five days a week. The quality of IPS can differ substantially across villages as, until recently, communes were required to establish IPS with their own funds. Consequently, IPS classes are often held in a teacher's home, in a community hall, or a pagoda (see Figure 2). IPS often lack the appropriate equipment, such as teaching and play materials or sanitary facilities. In most cases, the IPS lack even the most rudimentary equipment, such as tables and chairs.

To increase preschool access and to improve the unsatisfactory quality of IPS, the Cambodian government agreed to use the GPE II grant to establish 500 new formal community preschools. Some of these CPS replaced previously existing informal arrangements; others were established in villages that had no previous preschool or were too large to be serviced by one preschool alone. Conversely to IPS, a CPS benefits from uniform quality standards, such as a standardized building (see Figure 3), directly financed by the GPE II. CPS have a capacity of 25 children and are fully equipped with tables, chairs, a blackboard, and teaching materials. In partnership with the GPE representatives, MoEYS is responsible for the curriculum, recruitment, and training of teachers, as well as the monitoring of the running facility, including regular payment of teacher salaries. The CPS teacher is usually a 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.

[6] According to government data (MoEYS, 2017), out of 7,241 preschool facilities in Cambodia in 2016, 55% were SPS, 39% were IPS, and 6% were private preschools. However, these preschools are not evenly distributed across the country, and 38% of the 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 the GPE I.

2.3 Randomization and Data

The evaluation of the CPS program is based on a randomized controlled trial.[8] All sample villages are situated in the south and northeast parts of Cambodia, as the western part of the country was already covered by GPE I (see map in Figure 4). 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 is composed of 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 treatment group (120 villages), which received a CPS. An additional 127 villages received a CPS plus a demand-side intervention.[9] These demand-side interventions were in part implemented at the follow-up data collection in 2017, and hence, their impact will be evaluated on the basis of a later follow-up (2018) and discussed in a separate article.

Table 1 gives an overview of data collection activities and timing of the preschool construction. The analyses presented in this paper are based on two main waves of data collection: a baseline data collection in 2016 and a first follow-up in 2017.[10] Additionally, a brief monitoring survey was conducted in late 2016 to confirm that preschool construction proceeds 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 baseline survey and construction.

During the baseline data collection exercise in 2016, our survey firm sampled up

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

[9] 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, and 5 of them were duplicated or could not be identified after the randomization. Therefore, the total number of villages decreased to 305. We treated this drop-out as random and did not replace the villages.

[10] An additional follow-up is conducted in 2018 to determine impacts after two years and performance of the demand-side interventions. The findings will be published in a separate paper once the data becomes available.

[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.

to 26 eligible households per village.[12] Eligible households are composed of at least one child between 24 and 59 months old at baseline. Eligible children were therefore 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). Parent-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 measures five crucial domains of cognitive and physical child development: executive function, language, numeracy, and both fine- and gross-motor development.[14] Most child tests stem from the Measuring Early Learning Quality and Outcomes project (MELQO). MELQO tools were designed to provide a starting point for national-level adaptation of global measures of child development (see UNESCO (2017) for an overview) and have demonstrated adequate internal validity (Fernald et al., 2017; Berkes et al., 2018b).[15] Additionally, anthropometric measurements (height and weight) were 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 2–4 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, 7.9% for household attrition,

[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).

[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 Kampong Speu, Kandal, Prey Veng, Svay Rieng, and Takeo, the caregiver is often a grandparent. These are provinces with relatively high levels of manufacturing industry, 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. (2018b).

[15] Our version of the test is available on demand.

[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 at the age of 2–4.

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

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 sample villages. Attrition is slightly larger in the treatment group (+4.5%), yet the difference is not statistically significant.

Table 3 (household and caregiver characteristics) and Table 4 (child characteristics) show 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. One exception is preschool enrollment caused by the slight overlap of preschool construction and timing of the baseline survey (cf. Table 1). Children were 6.6pp more likely to be enrolled at preschool (last panel, Table 4). 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 control children, and since we do not measure any developmental difference between treatment and control at baseline, we consider the difference as negligible.

Table 3 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] 55% of households live on less than $100 per month. Caregivers, on average, have a formal education of 4 years. 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., 2018b), children belonging to 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) found 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 of development ahead from poorer children once they reach primary school age. Figure 5 summarizes wealth gradients in cognitive development, showing that while the gap widens between 3 and 5 in language and numeracy skills, things stay relatively constant for executive function and socio-emotional skills. Preschool is often thought of as a policy to mitigate these inequities. Below, we will show that the preschool program studied here falls short of these promises.

[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.

2.5 Preschool Quality

We use village survey data to show differences in quality measures between the types of preschools at baseline (Table 5) and follow-up (Table 6).[19] Table 5 documents that SPS are significantly different from CPS and IPS: SPS are larger (6 additional children when compared to an IPS, which serves around 21 students), and they have more material, such as chairs, tables, and blackboards. SPS 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 and have a significantly higher salary (on average, $90 per month against $35 for IPS teachers). Already at baseline, the quality of CPS appears better than that of IPS: CPS have better and 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 or training.

At follow-up (Table 6), SPS still offered a higher quality than CPS and IPS, but the CPS quality has further increased. CPS buildings are still reported to be larger and of better quality than IPS, but this time, CPS are reported to have more tables, chairs, and additional learning materials. Indeed by follow-up, almost all the CPS equipment had already been delivered. Yet once again, in terms of teacher quality, the difference in IPS and CPS is small, at least in the eyes of the village chief. Teachers in IPS and CPS seem to have seen their situation improved in a similar fashion: preschool teachers are more regularly and better paid at follow-up than they were at baseline. In all, Tables 5 and 6 indicate that the program has significantly improved the infrastructure quality of the community preschools: CPS have more materials and better premises. Yet the teaching quality – arguably the most important factor in early children's development – remains comparable in the eyes of the village chief.

3 Empirical Framework

As explained in the previous section, we evaluate the impact of the CPS in the context where alternative preschools (APS), i.e. SPS and IPS, are also available. The presence of close substitute preschools makes the identification and interpretation of some of the key treatment parameters ambiguous. In this section, we present the empirical framework and describe the strategies that we implement to identify the relevant treatment parameters. We start the section by presenting the assump-

[19] Note that the full study sample of all 305 villages is used in these tables to maximize statistical power. Since a CPS was also constructed in the two other treatment branches, and SPS and IPS are present, they can be used to document preschool quality. Note also that, as shown in Table 1, only a handful of CPS were already open at baseline, while almost all CPS were completed at follow-up. Hence, Table 6 is better suited to assess the final quality of CPS.

tions needed to identify the simple difference between the treatment and the control group (intention-to-treat or ITT) and the LATE, here called $LATE_{cps}$, i.e. the local average treatment effect of going to CPS. We then discuss how they are consistently estimated and how $LATE_{cps}$ can be decomposed into subLATEs. Finally, we present our strategy to bound the effect on home compliers and to pinpoint the subLATEs using additional assumptions.

3.1 Identification of ITT

We define $D_i \in \{c, a, h\}$ as capturing enrollment into CPS, into APS, or as not being enrolled in preschool (home care), respectively. Let Z_i be the instrumental variable that takes the value 1 for children in treatment villages and 0 otherwise. $D_i(Z_i)$ and $Y(D_i)$ are potential enrollment and outcome variables.

We rely on the independence of Z to identify the ITT. As mentioned, Tables 3 and 4 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[20]. While a few treatment group villages are in the vicinity of control group villages, we have no reason to believe that the construction of 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 their recruitment is therefore not related to the availability of teachers in the control group.

Under A1 and A2, we estimate ITT effects using the following regression:

$$Y_{iv} = \alpha + \beta Z_{iv} + \boldsymbol{W_i}\boldsymbol{\gamma} + \mu_v + \epsilon_{iv} \tag{1}$$

where Y_{iv} denotes the outcome of child i and village v. μ_v is a village-specific error term component, assumed to be uncorrelated with W and Z, and ε_{iv} is an unobserved within-village error component. We use standard errors clustered at the village level to account for the randomization implemented at the village level. W is a set of control variables used to improve precision and to correct for finite-sample imbalances: it includes (i) all children level baseline test scores; (ii) province fixed effects; and (iii) child age and gender.[21] We will present results with and without covariates, but the conditional regressions will be our preferred estimation, since

[20]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.

[21]We replace missing test scores by the sample mean and interact them with a dichotomous variable indicating a missing value. Age is measured as a trimester fixed effect and also imputed if missing.

controlling for them considerably reduces the residual variation. The outcomes of interest Y, specified in the pre-analysis plan (Berkes et al., 2017), include (i) the school construction collected at the village level (see Section 4.1.1); (ii) the enrollment in, and month of exposure to, each childcare arrangement (see Section 4.1.2); and (iii) the children's cognitive and socio-emotional performance and the parental response to the school construction (see Section 4.1.3).

3.2 Identification of $LATE_{cps}$

To measure the treatment effect on CPS compliers ($LATE_{cps}$), we reformulate the assumptions of the traditional LATE framework by Angrist et al. (1996), where D is generally a binary variable to adapt them to our setting where D takes three values (c,a,h).

In addition to assumptions A1 and A2, the identification of $LATE_{cps}$ relies on the first-stage assumption (A3), which we deem to fulfill in our setting due to the high CPS take-up of CPS (see supra). In the presence of close substitute, we reformulate the exclusion restriction in the following way:

Assumption 4 - *Exclusion Restriction (A4)*

(i) $Y_i(c,1) = Y_i(c,0)$,

(ii) $Y_i(a,1) = Y_i(a,0)$,

(iii) $Y_i(h,1) = Y_i(h,0)$.

Case A4 (i) refers to the typical John Henry or Hawthorne effect violation. In our context, such a violation seems unlikely, since children cannot easily or willingly manipulate their test scores. Furthermore, adult participants were not informed that they are part of an experiment with treatment and control groups. Case A4 (ii) is violated if CPS construction modifies the learning opportunities in APS. For instance, the construction of a CPS may either decrease class size or modify peer composition in the remaining APS, with the likely consequence to improve the learning conditions of a-never-takers[22]. Finally, case A4 (iii) may be violated if the CPS had an effect on non-compliers' parents; for instance, through a higher level of involvement at home. We discuss more at length the validity of the exclusion restriction in Appendix C, Table C.1, and Table C.2, and conclude that assumption A4 is very likely to hold in our context.

While A4 (i) - (iii) are necessary conditions for the $LATE_{cps}$ to be identified, they are not sufficient without a monotonicity assumption that takes close substitutes into account:

[22]Since APS are essentially composed of SPS, presumed to be of better quality (see supra), we hypothesize that if there is a composition effect, it would entail a more favorable peer composition in APS. The reverse is not impossible, though.

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

(i) ch-defiers: $D_i(0) = c$, $D_i(1) = h$,

(ii) ca-defiers: $D_i(0) = c$, $D_i(1) = a$,

(iii) ah-defiers: $D_i(0) = a$, $D_i(1) = h$,

(iv) ha-defiers: $D_i(0) = h$, $D_i(1) = a$.

Cases A5 (i) and (ii) together are analogous to defiers in the traditional LATE framework. Since enrollment into CPS is zero in our control group, we can rule out these two cases and consider the traditional monotonicity assumption (Angrist et al., 1996) as fulfilled.

Yet, the cases (iii) and (iv) are theoretically possible. They correspond to the opposite situation, where the CPS construction would either increase (ha-defiers) or decrease (ah-defiers) the APS attendance of children who would have stayed at home otherwise. While the existence of ah-defiers is very unlikely, as the construction of CPS is unlikely to reduce the overall demand for preschool, ha-defiers deserve more attention.[23] 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 lower 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 APS are already at capacity when the CPS opens, children enrolling in CPS would make room for ha-defiers. This situation would again violate A5.[24]

In our context, however, presence of ha-defiers is unlikely. First, as pointed out in Table 7, 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. Second – and perhaps more importantly – in the vast majority of cases, the construction of a CPS caused the IPS to shut down: only 7 IPS remained open after the 103 CPS were constructed, for a total of only sixty-six children enrolled in an IPS when a CPS is built. In the vast majority of cases, children staying at home could simply not have enrolled into an APS because the APS did not exist anymore.

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

[24] 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.

In all, APS and CPS were not rationed, and in most cases, APS shut down after the CPS was constructed. Therefore, although ha-defiers are likely to be a concern in many settings (especially in high-income countries), in our context, we can rule out the presence of ha-defiers and assume A5 valid.

Figure 6 provides a visual representation of the different sub-populations in our sample under the stated assumptions. In the absence of the intervention, i.e. the control group scenario (left panel), CPS compliers would have either stayed home or attended an APS. After the CPS construction occurred in the treatment group (right panel), children in the treatment group give rise to four principal strata (Feller et al., 2016):

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): $D_i(0) = h$, $D_i(1) = c$,

By virtue of A3, the CPS construction does not affect the never-takers: a-never-takers stay enrolled in APS irrespective of the treatment assignment, while h-never-takers remain at home, and their respective outcomes are therefore independent of Z. Only compliers in the treatment group are positively affected by the treatment. Yet, $LATE_{cps}$ captures two effects: the effect on a-compliers ($LATE_{ac}$) and the effect on the h-compliers ($LATE_{hc}$).

The fact that $D_i(0)$ takes two different values for compliers switching to CPS is of particular importance here. While it is usually assumed that the counterfactual to take-up is homogeneous, in a context like ours, $D_i(0)$ can be either a or h. This creates a form of unobserved heterogeneity, as $D_i(0)$ are not directly observable. Consequently, if $D(0) \not\perp \boldsymbol{W}$, then any heterogeneous treatment analysis using $\boldsymbol{W}$ would confound heterogeneity by *observable characteristics* with heterogeneity by *counterfactual substitute*. In the following, we describe how to take close substitutes better into account.

3.3 Identifying Substitution and Characterizing Principal Strata

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

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

In presence of a close substitute, the $LATE_{cps}$ can be further decomposed into two subLATEs (Kline and Walters, 2016):

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

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):

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

Figure 6 provides a visual representation of the parameters in equation (2): 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 subLATEs, $LATE_{ac}$ and $LATE_{hc}$.

To conclude our empirical framework, we provide in Figure 7 a similar chart than before, but this time with a modification in notation to better describe the principal strata. Let $Y_p^{z,t}$ be the expected outcome variable (typically here the performance of the child at a development test) at time t (taking 0 for baseline, 1 for follow-up), in experimental branch z (1 for treatment, 0 for control) and for principal strata p. p takes the four values: a-compliers (ac), h-compliers (hc), a-never-takers (ant), and h-never-takers (hnt). Hence:

$$\begin{aligned}
Y_{ac}^{z,t} &= E[Y^t(z)|D(1) = c, D(0) = a] \\
Y_{hc}^{z,t} &= E[Y^t(z)|D(1) = a, D(0) = h] \\
Y_{ant}^{z,t} &= E[Y^t(z)|D(1) = a, D(0) = a] \\
Y_{hnt}^{z,t} &= E[Y^t(z)|D(1) = h, D(0) = h]
\end{aligned}$$

To summarize, Table 8 maps the different principal strata for each value of Z and of D as well as their expected value notation under A5. Using this notation and Figure 7, each subLATE can be written:

$$\begin{aligned}
LATE_{ac} &= Y_{ac}^{1,1} - Y_{ac}^{0,1} \\
LATE_{hc} &= Y_{hc}^{1,1} - Y_{hc}^{0,1} \\
LATE_{ant} &= Y_{ant}^{1,1} - Y_{ant}^{0,1} = 0 \\
LATE_{hnt} &= Y_{hnt}^{1,1} - Y_{hnt}^{0,1} = 0
\end{aligned} \tag{4}$$

again with the effects on non-compliers equals to zero. Similar to Feller et al. (2016), we also introduce $A_d^{z,t}$, the weighted average of $Y_p^{z,t}$ for the children staying at home

in the control, going to APS in the control or going to CPS in the treatment:

$$\begin{aligned} A_c^{1,t} &= S_{ac} * Y_{ac}^{1,t} + (1 - S_{ac}) * Y_{hc}^{1,t} \\ A_a^{0,t} &= S_{ant} * Y_{ant}^{0,t} + (1 - S_{ant}) * Y_{ac}^{0,t} \\ A_h^{0,t} &= S_{hnt} * Y_{hnt}^{0,t} + (1 - S_{hnt}) * Y_{hc}^{0,t} \end{aligned} \tag{5}$$

with S_{hnt} (resp. S_{ant}) the share of h-never-takers (a-never-takers) defined analogous to (3).[25] Note that (5) can be used to describe principal strata at follow-up but also at baseline. Importantly, when $t = 1$ in equations (5) and (4), all parameters are identified except for $Y_{ac}^{1,1}$ and $Y_{hc}^{1,1}$.[26]

3.4 Bounding $LATE_{hc}$

Equation (2) makes explicit the challenges faced by researchers when estimating the impact of a policy in a context of close substitutes. Under A1–A5, $LATE_{cps}$ is identified, but its interpretation is ambiguous. The magnitude of $LATE_{cps}$ crucially depends on the quality of the counterfactual APS and the share of a-compliers. $LATE_{cps}$ is a weighted average of two subLATE parameters, which are not identified without further assumptions. As mentioned in the introduction, due to its potential for high external validity, we consider the $LATE_{hc}$ of particular high interest in the preschool literature.

Additional observations and assumptions are needed to derive bounds for $LATE_{hc}$. From equation (2), one particularly appealing assumption is the following:

Assumption 6 - *For every type of compliers (a and h), preschool is never detrimental, and its impact increases with resources (A6):*

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

The first inequality simply implies that for a-compliers, switching from APS to CPS is not detrimental on average, i.e. $0 \leq LATE_{ac}$ or $Y_{ac}^{1,1} \geq Y_{ac}^{0,1}$. As we will show further below, all or almost all a-compliers are coming from IPS, as opposed to SPS. Given the resources devoted to CPS in comparison with IPS, as discussed in Section 2.5, we believe that the left side of the A6 inequality is a very likely assumption. The second inequality implies that the h-compliers benefit more from CPS than a-compliers, i.e. $Y_{ac}^{1,1} - Y_{ac}^{0,1} \geq Y_{hc}^{1,1} - Y_{hc}^{0,1}$. Intuitively, it remains a light assumption in our context. Since a-compliers would have benefited from a

[25] $S_{hnt} = \frac{P(D=h|Z=0)-P(D=h|Z=1)}{P(D=h|Z=1)-P(D=h|Z=0)}$ and $S_{ant} = \frac{P(D=a|Z=0)-P(D=a|Z=1)}{P(D=a|Z=1)-P(D=a|Z=0)}$.

[26] $Y_{ac}^{0,1}$ and $Y_{hc}^{0,1}$ are non-parametrically identified:

$$\begin{aligned} Y_{hc}^{0,1} &= (A_h^{0,1} - S_{hnt} * Y_{hnt}^{0,1})/(1 - S_{hnt}) \\ Y_{ac}^{0,1} &= (A_a^{0,1} - S_{ant} * Y_{ant}^{0,1})/(1 - S_{ant}) \end{aligned}$$

worse but still comparable form of childcare (mostly IPS) in the absence of the treatment, it is very likely that h-compliers, who benefited only from the care of their caregiver, would benefit more from a CPS enrollment than a-compliers. Yet, A6 is not guaranteed. If h-compliers were to benefit from a particularly stimulating home environment, e.g. due to parents who heavily invest in their children's education at home, then A6 might be violated. Similarly, if going to an APS is detrimental to the a-compliers, e.g. if APS provides a much poorer environment than the home environment of a-compliers, it is not impossible that switching from APS to CPS would be more valuable for the a-compliers than transitioning from home to CPS. In both cases, violation of the right side of the A6's inequality would require that the h-compliers' home environment would be much more favorable than the one provided by a-compliers' parents and/or that the a-compliers would perform much worse than h-compliers in the absence of the treatment.

By definition, a-compliers are always going to preschool, regardless of their treatment status: they can therefore be expected to have parents who are more educated, more concerned by education, more aware of the virtue of preschool, and possibly able to provide a more stimulating environment at home. Hence, we would expect the performance of untreated a-compliers to be higher than the performance of h-compliers. Using equation (5), we can describe and compare $Y_{hc}^{0,1}$ and $Y_{ac}^{0,1}$. Although we sometimes lack statistical power, Table 9 overall confirms our expectations: parents of a-compliers are performing significantly better in our cognitive parenting measure than parents of h-compliers. Further, while mostly insignificant, child test scores are consistently lower among h-compliers than among a-compliers. Hence, we conclude that a-compliers would most likely benefit from a better home environment than h-compliers. We interpret this as evidence in favor of A6: h-compliers in our sample benefit more from CPS enrollment than a-compliers.

Additional suggestive evidence for A6 can also be inferred from the preschool literature. By definition, a-compliers are most likely to be enrolled at any preschool. Even though h-compliers have enrolled in preschool after receiving the CPS, they would not have done so without the provided supply-side intervention. We interpret this as supporting evidence that h-compliers are characterized by a relatively lower propensity to enroll at preschool. While in a classical Roy model, the individuals with the highest return to a program are expected to be the first to take it up, negative selection on return is the more common finding in the preschool literature (or "inverse Roy model"). Kline and Walters (2016), Cornelissen et al. (2018), or even Bouguen et al. (2018) have already reported that children with the lowest return to preschool are the most likely to participate. The wealthier and more educated households, who are likely to provide more valuable parenting, are more likely to send their children to school than poor families, despite the fact that children from poor families may be the ones with the highest return to preschool. Two explanations

can account for such behavior. First, wealthier and more educated families may have a higher perceived return to preschool: they are more aware of the value of education, and hence sensitive to any new education technologies made available to them. Second, the caregivers from wealthier backgrounds are likely to have more valuable outside options (higher wages or higher return to leisure), making them more prone to avail themselves from services that may relax their time constraints. This evidence from the preschool literature, combined with the preschool quality data, substitution patterns described further below, and Table 9, all convince us that A6 is very likely to hold and is, in fact, a very plausible assumption.

Under A1-A6, 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 benefit equally from the new intervention on average. Hence, $LATE_{cps}$ is the local average treatment effect of both sub-populations. Under the high bound, we assume:

$$\begin{aligned} LATE_{ac} = 0 \iff LATE_{h}^{HB} =& \frac{LATE_{cps}}{(1 - S_{ac})} \\ =& \frac{ITT}{TU_c * (1 - S_{ac})} = \frac{ITT}{TU_c * (1 + \frac{TU_a}{TU_c})} \\ =& \frac{ITT}{TU_c + TU_a} = \frac{ITT}{TU_{ps}} \equiv LATE_{ps} \end{aligned} \tag{7}$$

with $LATE_{ps}$ the effect of *any* preschool enrollment instrumented by Z. In other words, if the outcome of a-compliers was not affected by the treatment, estimating the traditional LATE with D as enrollment into any preschool would not constitute a violation of the Angrist et al. (1996) exclusion restriction and identify $LATE_{hc}$. Essentially, our arguments imply:

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

$LATE_{hc}$ is very likely to be bounded by $LATE_{cps}$ and $LATE_{ps}$. As described in the next section, additional and more stringent assumptions allow us to obtain a more precise estimate of $LATE_{hc}$.

3.5 Estimating $LATE_{hc}$ Using Conditional LATE

Although the bounds are our preferred estimation strategy for $LATE_{hc}$, as they rely on a less restrictive set of assumptions, we implement a conditional LATE strategy to pinpoint where the subLATE parameters are most likely to lie.

Using additional baseline characteristics as instruments and under a constant subLATE assumption, we can isolate $LATE_{hc}$ and $LATE_{ac}$. Kline and Walters (2016) – see also Hull (2018) and Feller et al. (2016)[27] – show that subLATEs can be identified by interacting the randomly assigned preschool construction program with observed covariates. For example, let $X_i \in \{0, 1\}$ be a dummy variable strongly correlated with the APS enrollment (for instance, presence of APS in the village at baseline). For each value of X, the conditional $LATE$ can be written as:

$$\begin{aligned} LATE_{cps}(1) &= S_a(1)LATE_{ac} + (1 - S_a(1))LATE_{hc} \\ LATE_{cps}(0) &= S_a(0)LATE_{ac} + (1 - S_a(0))LATE_{hc} \end{aligned} \tag{8}$$

As we can estimate $LATE_{cps}(X)$ and $S_a(X)$ for each value of X, we can identify the two unknown subLATEs by solving the system of equations. As it is apparent in (8), the identification relies on the assumption that $LATE_{ac}$ and $LATE_{hc}$ do not depend on X; i.e. subLATEs are constant on X. In other words, the variation of the $LATE_{cps}(X)$ derives entirely from variation in $S_a(X)$, while the subLATEs remain constant on X.

As demonstrated by Feller et al. (2016) and Kline and Walters (2016), (8) can be estimated using a 2SLS, with $\mathbb{1}_{\{D_i=c\}}$ (enrollment in CPS) and $\mathbb{1}_{\{D_i=a\}}$ (enrollment in APS) as endogenous variables and Z and $\boldsymbol{X} * Z$ as instrumental variables. The model takes the following form:

$$Y_i = \beta_0 + \beta_1 \ \mathbb{1}_{\{D_i=c\}} + \beta_2 \mathbb{1}_{\{D_i=a\}} + \beta_3 \boldsymbol{X}_i + \beta_3 \boldsymbol{W}_i + u_i \tag{9}$$

where Y_i is a follow-up outcome, $\boldsymbol{X}$ a set of additional instruments, and $\boldsymbol{W}$ additional control variables (here, age and gender). β_1 captures the $LATE_{hc}$, and β_2 captures the effect of going to APS. To derive the $LATE_{ac}$, we therefore subtract β_2 from β_1. $\mathbb{1}_{\{D_i=c\}}$ and $\mathbb{1}_{\{D_i=a\}}$ are both endogenous and instrumented by:

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

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, in this context, does not mean that the treatment effect is constant among all sampled individuals. It means that among compliers, if the treatment is heterogeneous, such heterogeneity should be *between* a- and h-compliers groups and not *within* them. In other words, while heterogeneity among compliers is possible, the treatment heterogeneity must be

[27] Previously, a very similar strategy was used in another context by Banerjee et al. (2007).

driven by the heterogeneity of *counterfactual substitute* rather than by the standard heterogeneity of *observed characteristics* (see infra section 3.2). Note that when X is composed of several variables, the validity of this assumption can be tested using an over-identification test.

One of the practical issues of (9) is that the set of X variable needs to be sufficiently predictive of the substitution behavior to secure a good first stage. In our context, we include village and infrastructure indicators in a first specification, and we add individual level characteristics in a second.[28] These variables were chosen for their ability to predict the share of a-compliers in each village. We will provide p-values from over-identification tests below to assess validity of the constant subLATE assumption for each set of $\boldsymbol{X}$ variables.

4 Results

4.1 Reduced-Form Estimates

4.1.1 School Construction

We begin our empirical analysis by documenting in Table 10 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] The fact that none of the control group villages received a CPS confirms that the Cambodian government strictly respected the study protocol. Less anticipated was the number of other preschools available in both treatment and control villages. 81% of the control group villages had at least some kind of preschool, and hence, our treatment increases availability of any preschool by just 12 pp.

Furthermore, CPS availability declines by a significant 55 pp in the treatment group, while the availability of state preschools (SPS) remains approximately unaffected (insignificant -6 pp). As confirmed by our field visits, IPS were often shut down or were turned into CPS as soon as the new preschool building became available. Conversely, the SPS, which is already a formalized form of preschool, remained available to the children.[30] Table 10 confirms that most of the substitution occurred

[28] Village level indicators include a dummy for above median village population size, as well as land area, baseline presence of a primary school, baseline presence of a secondary school, and province fixed effects. Individual level characteristics include baseline caregiver raven score, caregiver education, household poverty dummy, and baseline child test scores.

[29] As a comparison, in Bouguen et al. (2018), differential take-up at the village level was 43 pp. Martinez et al. (2017), and in Indonesia (Brinkman et al., 2017), all control villages received the program by midline. 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 Gambia.

[30] In villages where both CPS and SPS were available, five-year-olds would often register at SPS,

between IPS and CPS and that children enrolling in SPS have been almost unaffected by the program. This substitution pattern has important implications for the interpretation of our results: Since IPS are arguably of much lower quality than SPS (see Tables 5 and 6), the fact that a-compliers would have enrolled in IPS if assigned to the control means that a-compliers are likely to contribute positively to the overall treatment effect. Although in the rest of the analysis, we will still consider the substitution pattern to exist between CPS and APS, the reader should keep in mind that the vast majority of the substitution is actually occurring between IPS and CPS.

4.1.2 Preschool Enrollment

To study the enrollment patterns at the child level, we explore in Table 11, the enrollment and exposure of children, separately by preschool type. Assignment to the treatment group had a significant effect of about 39 pp on CPS enrollment (47 pp for the five-year-olds). Since 14% of treatment group villages did not receive a CPS until follow-up, the CPS take-up rate in villages with CPS is about 45 pp (55 pp for five-year-olds) on average. Such level of (differential) enrollment is in line with our most optimistic scenario of the power calculations.[31] It is also higher than the enrollment rates reported in low-income countries elsewhere in the literature.[32] Finally, such high take-up rate confirms the relatively high level of Cambodian parents' interest in preschool education.

We explore the potential reasons for non-enrollment in Table 7.[33] The most common reasons relate either to the self-sufficiency and emotional maturity of the child (the child is too young; too active; is afraid; does not speak well enough; or refuses to go) or to various practical reasons (afraid child gets hurt on way to school; no one there to pick up child; school is too far away). Preschool quality (inadequate school facility, unreliable or unqualified teacher) are much less often evoked while financial constrains are almost never mentioned as reason for non-enrollment. Note finally that since classes are typically given between 7 and 9 am, preschools are unlikely to relax labor supply constraints of the mother and may even constitute an additional constraint when the child is enrolled.

Caregiver also reported the enrollment history of each child. Exposure to CPS led to an increase of about 3.4 months, while it decreased by about 2.4 months in

while three- to four-year-old children would register at CPS.

[31]The most optimistic scenario was 51.43%.

[32]Martinez et al. (2017) reported a differential take-up of about 33 pp, 24 pp in Indonesia (Brinkman et al., 2017), 9 pp in Gambia (Jung and Hasan, 2016), and 25pp in the previous preschool impact evaluation in Cambodia (Bouguen et al., 2018).

[33]Eliciting reasons for non-enrollment is always challenging: it is often multi-factorial, affected by social desirability, and influenced by the way the question is framed. We should therefore take these descriptive statistics with caution. Here, we first asked which reasons for non-enrollment apply among a list of possible reasons, including the possibility to add an additional reason; second, we let caregivers rank the most important to third most important reasons.

APS, for an overall increase of about 1 month. This is an important result when interpreting the magnitude of the ITT effects on test scores: even when a-compliers are taken into account, results are driven by a short period of preschool attendance. While statistical power is sufficient to detect even small impact, we cannot expect the ITT results to be very large.

Finally, Table 11 provides children level information about the substitution patterns. As seen at the village level in Table 10, assignment to treatment negatively affects the probability to attend APS as well as the APS exposure. As in Table 10, 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. Besides, Table 11 provides all the necessary information to calculate S_{ac} as in equation (3). Under the extended monotonicity assumption (A5), 28.3% of the sample are a-compliers. Hence, among the 38.9% of the sample who complied, the share of a-compliers is 73%, while 27% of the compliers would have stayed at home if the CPS construction had not occurred. Similarly, 70% of the overall exposure in terms of months enrolled is coming from the a-compliers, while 30% is coming from the h-compliers. Note that despite higher level of CPS enrollment, the five-year-olds do not drive up the share of a-compliers: they participate more in CPS, but they also substitute more from APS to CPS (S^5_{ac}=80%).

4.1.3 ITT Impact on Child Performance and Parental Response

We assess the ITT impact of the CPS construction in Table 12. As indicated in equation (1), we present both the treatment-control differences (column 1) and the treatment coefficients controlled for baseline characteristics (column 2). 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[34] and the outcomes[35] variables we use, column (2) is our preferred estimate.

ITT results point towards a positive effect of the CPS construction on the performance of children. Children in treatment villages perform about 0.05 to 0.06 SD higher in treatment than those in control villages in cognitive test scores, but the

[34]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 includes: child gender, child age (trimester fixed effects), province fixed effect, and all baseline child performance measures (test scores). We are therefore more conservative than the PAP as we do not include any household characteristics. Since household characteristics were very well balanced, none of the them made any significant improvement in term of precision and none of them modified our results significantly.

[35]In the PAP, we included the gross and fine motor skills together and did not mention the anthropometric measures (we were unsure whether anthropometric measures would be collected). Given the low level of correlation between anthropometric measures, gross motor, and fine motor items, we decided in Table 12 to regroup the anthropometric measures and the gross motor test score in the physical development test score. Fine motor skills, which are a prerequisite for some dimensions of literacy skills, such as writing letters, were included in the cognitive measures.

difference remains insignificant for physical development. To increase precision and alleviate the issue of multi-hypothesis testing, we aggregate all the cognitive and physical development test scores into two indexes. The effect on the cognitive index becomes more significant and is evaluated at 0.051 SD, while the physical development index is not significant. The cognitive effect corresponds to approximately a tenth of the initial quintile cognitive gap, or tenth of a year of child development between three and five (Berkes et al., 2018b). Given the large substitution pattern documented previously, the small magnitude of the CPS impact should be interpreted with care, as it is driven by both a and h compliers that may have very different treatment effects. We come back to that point in the following section.

Table 12 also indicates that results are driven by the five-year-olds, with an impact that reaches 0.1 SD (aggregate measure of cognition). The fact that five-year-olds are more likely to enroll in CPS and have a larger share of a-compliers, yet have a similar effect on enrollment into preschool in general (see Table 12), seems to suggest that the a-compliers also contribute to the overall effect – in other words, switching from APS to CPS is actually valuable to the child cognitive performance, a conclusion that we already touch upon when looking at the substitution patterns. Another explanation is that the CPS curriculum is more adjusted to the 5-year-olds, and hence, they are the ones benefiting most from the CPS. Again, we will come back to this point in the next section.

We complete the analysis of the ITT results by reporting the performance of parents in Table 13. Table 13 covers two important aspects of parents' attitude towards education: their involvement at home and their perception of the benefit of formalized education. Most of these measures are self-reported and should therefore be analyzed with care.

The first important finding is that we do not find any evidence of a substitution between parental involvement and preschool. If anything, CPS availability in the village positively impacts cognitive parenting. Since cognitive parenting[36] is also the parental dimension with the strongest partial correlation with children's cognitive development tests, preschool appears to be a complement to parenting. Yet, as documented in Berkes et al. (2018b), cognitive parenting is unlikely to entirely explain the ITT cognitive effect on children: we estimate that a 1 SD effect on cognitive parenting has a 0.1 SD effect on children's cognitive performance – a tenth of the total effect size on children. The complementarity – or the absence of substitution – between parents and preschool, also documented elsewhere – see, for instance, Ozler et al. (2018) – suggests that investing in preschool in a village may stress the importance of education and highlight the role that parents may play in the child's

[36]The cognitive parenting score measures how parents actively interact with the child in ways that are likely to develop cognitive competencies: games, reading books and playing with toys or objects (Berkes et al., 2018b).

development.[37]

This interpretation is confirmed by the second panel of Table 13: all the self-declared measures point to more favorable parental attitudes towards education in the treatment group. Parents are more inclined to send their children to school earlier, and they report a higher perceived return to primary and secondary education. While the perceived return to education in the control group is not so far away from what the latest research on return to education suggests – 5.5% in primary school in the control group against the 6.6% estimated by Lall and Sakellariou (2010) in Cambodia – families in the treatment group evaluate the primary school's return at 6.1 % per year, almost exactly like the estimates provided by Lall and Sakellariou (2010). Similarly, in secondary school, the perceived return increases from 6.2% in control to 6.7% in treatment. Here again, the construction of school seems to have affected the mindset of parents towards a higher perceived return to education.

4.2 Estimating $LATE_{hc}$

Although small, the magnitude of ITT effects should be interpreted with care: ITT effects reflect the overall take-up rate of CPS and the effect of two sets of compliers that are likely to have performed differently in CPS. As mentioned in the introduction, we hence focus on the identification of $LATE_{hc}$, which remain an important and understudied parameter of the literature. We start our investigation using a bounding strategy. As shown in Section 3, the bounds, rely on a set of relatively light assumptions. Then, we will use several instruments to pinpoint the $LATE_{hc}$.

4.2.1 Bounding $LATE_{hc}$

We present our estimation of the bounds in Table 14. While the low bound is equivalent to an estimate of $LATEcps$, the high bound is equivalent to $LATEps$, as defined in Section 3.4. We start the analysis by looking at the effect on month of exposure, here measured as any school exposure since birth. Bounds 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 high bound seems much more likely. Indeed, children who switch from APS to CPS are unlikely to experience a different level of preschool exposure. As a result, at least for exposure, we can assume $LATE_{ac} = 0$, which corresponds to the high bound in equation (7). Consequently, the h-compliers are very likely to have spent 9.5 month at school in total. Still according to Table 14, 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. Here again, the five-year-olds drive the impact, with a high bound on the aggregate measure as high as 0.99 SD. Although

[37] If the effect is very large and also affects parents that did not enroll their child into CPS, this would be a violation of assumption A4. We discuss this in Appendix C.

the high bound seems large, it is not out of proportion when compared with the few studies that have measured the effect of a school construction on children who would have stayed at home otherwise.[38] Besides, CPS attendance among h-compliers does not significantly impact parenting measures. We are therefore incline to think that parenting skills is a substitute to preschool enrollment, as the results are lining toward the positive side.

Our results therefore confirm the role that preschool plays in improving children's cognitive performance. The results so far are consistent with an impact as high as 0.45 SD. These results would mean that the cognitive gap, measured at baseline to be of about 0.6 SD, could be significantly reduced if targeted to the right children. The results show clearly that low ITT effects are not inconsistent with substantial effects on home compliers. A finding that is an important contribution to the preschool literature where low ITT effects were often interpreted as evidence against sizable effects of preschool in general.

Although the bounds are our preferred measure of impact for the $LATE_{hc}$, being able to pinpoint the exact location of the parameter would help provide a better understanding of the preschool impact.

4.2.2 Beyond the Bounds

Under the constant subLATE assumptions outlined in Section 3.5, we use additional instrumental variables interacted with Z to estimate the subLATE effects as in equation 9. In Table 15, we provide a comparison between our bounds and the subLATEs for two different sets of additional instruments. The two first columns provide the bounds as in Table 14, with estimates for the $LATE_{hc}$ varying between 0.13 and 0.45 SD. We then estimate the subLATEs using first, a set of village level characteristics and province dummies and second, our preferred estimation where we add caregiver and child level baseline covariates to the instruments. Although the estimates remain imprecise (essentially due to the small share of h-compliers), results are generally in line with our estimation of the bounds. For instance, while the effect on language and numeracy is bounded between about 0.12 SD and 0.44 SD, our preferred specification for these two dimensions suggests that the $LATE_{hc}$ effect is above 0.2 SD, i.e. almost twice as high as the low-bound but much lower than the higher bound. For all outcomes and both specifications, the $LATE_{hc}$ is significantly different from the upper bound, yet never significantly different from the lower bound. Further, the $LATE_{ac}$ is located at a much lower level (around 0.08-0.09 SD for language and numeracy). While the point estimates suggest that the CPS program has improved the quality of the preschool for a-compliers, the $LATE_{ac}$ is not significantly different from zero for language and numeracy. While

[38] 1.3 SD is reported by Burde and Linden (2013) in Pakistan, and 2.2 SD by Kazianga et al. (2013) in Burkina Faso, both for a day shift in primary school.

the pattern is similar for the executive function measure, we should interpret this parameter with care due to a rejected over-identification test.

The effects on fine motor and socio-emotional skills are not entirely in line with our previous assumptions. For both outcomes, our LATE(X) results indicate a violation of assumption A6 since $LATE_{ac}$ is estimated to be above $LATE_{hc}$. 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 the APS is particularly ineffective and potentially detrimental to the child development in term of socio-emotional and fine motor. This finding resonates with a few empirical evidence. Enrolling children in preschool too early is sometimes suspected to have negative effects on socio-emotional skills (Baker et al., 2008). It is hence possible, in our context, that APS were so poorly equipped in material and building that they affected negatively socio-emotional performance. As a result, $LATE_{ac}$ is large and (almost) significant while better equipped preschools like CPS, which offer a more satisfactory education environment, are less likely to affect negatively the socio-emotional development of a child. Similarly, the $LATE_{ac}$ effect on fine motor skills is positive and significant (0.18 SD) while the CPS effect is small may also be interpreted as a negative effect of the APS. As shown in Table 5 and 6 the main difference between IPS and CPS is in material and the infrastructure: IPS 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 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.

Lastly, conversely to Kline and Walters (2016), our results have strong first stages (F-test above 6) and fails to reject the over-identification test, in most cases (except for our preferred specification of the executive functions). It suggests that, the homogeneity assumption – i.e. the subLATE does not vary on $\boldsymbol{X}$ – is valid in our context. This is true for all outcomes except for executive function where the over-identification test rejects the null hypothesis. For that specific dimension, we would hence prefer to refer to the bounds that are based on a lighter set of assumptions.

All in all, although the estimated subLATEs are only valid under a constant treatment effect assumption and are fairly imprecisely estimated, they provide an idea of where the true parameters are likely to be located. The LATE(X) estimates are consistently closer to the low bound and entirely inconsistent with the high bound. Taken at face value, Table 15 suggests that for 9.5 months of exposure (about one school year), children who would have stayed at home in absence of CPS perform, on average, about 0.20 SD better on the summary index when enrolled in CPS. If targeted to the right compliers, such a program could therefore reduce by about a third the total cognitive inequality measured at baseline.

4.3 Does Preschool Really Prevent the Cognitive Gap?

Thus far, we only touched upon the central question of cognitive inequality. Yet substantial cognitive differences between children from different socioeconomic backgrounds are among the main justifications for implementing early childcare programs (Heckman, 2006). Preschools, together with other ECD interventions (nutrition, cognitive stimulation), are expected to close the cognitive gap between children from different socioeconomic backgrounds, a gap that will be very difficult to close at later ages. In Cambodia, we have already mentioned that the numeracy and languages gap is widening during the three- to five year old window (see Figure 5). Implementing a CPS during that specific time period appears to be exactly what is needed to reduce the cognitive gap, at least for both of these dimensions.[39]

Unfortunately, we already have reasons to believe that the reduction of cognitive inequality is unlikely to happen in our context. Using the principal strata description strategy presented earlier, Table 9 shows that the compliers (a or h taken together) belong to a sort of middle class in terms of wealth and parenting measures in our sample: they are typically less educated and wealthier than a-never-takers, but are significantly wealthier and more educated than the h-never-takers, who represent a large group of deprived parents (44% of the sample according to Table 11. Since the a and h compliers are composed of middle-class parents, little is to be expected in terms of inequality reduction. At best, inequality should be entirely unaffected by the CPS construction. We verify this intuition by estimating the (ITT) heterogeneous treatment effect in Table 16. We use the baseline wealth index used in Figure 5, comparing the cognitive performance at follow-up of the quintile 1 (Q1) versus the quintile 5 (Q5). The results confirm our intuition. According to Table 16, the cognitive gap has not been closed; if anything, it might even have been increased, with Q5 treatment effect significantly outperforming Q1 effect on numeracy skills. Note that the difference between Q1 and Q5 is driven by both a larger CPS attendance among Q5 and a larger substitution rate (S_{ac}) among Q1: wealthier children are more likely to go to CPS but are less likely to have gone to APS in absence of the program. Given that we expect that preschool's impact is constant across X, we interpret Table 16 as evidence of a poor targeting of the CPS program. A better targeting of the CPS to the low achievers could (easily) restore the progressive nature of preschool. Cash transfers, nudges, free lunch or transportation could be implement to foster the performance of the CPS program. We look more specifically at ways to improve targeting in a companion paper (Berkes et al., 2018a).

[39]To be sure, one could claim that for stunting and executive function, the window of opportunity has already passed, as the gap is already 0.6 SD for three-year-olds in executive function and substantial for stunting. It seems pretty clear that preschool is not instrumental for these dimensions and should be complemented at early ages with nutrition, health, or early stimulation programs.

5 Conclusion

We analyze in this article 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 fundamentally different type 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 term 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-population, which is implicitly what standard treatment parameters (ITT and traditional LATE) produces, 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 (subLATEs) is necessary to produce more external valid statements and make appropriate policy recommendations.

We rely on a large and well-implemented preschool construction program, evaluated using a Randomized Controlled Trial, to produce three important results. First, using the traditional reduced-form tools (here ITT), we show that the preschool construction program increases preschool attendance (here by +39 pp) and improves the performance of three- to five-year-old children (+0.05 SD). Five-year-olds are more likely to enroll (+45 pp) and benefit more from the CPS (+0.11 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 previous articles studying preschool impact reported similar substitution patterns. In this paper, we suggest that in presence of close substitution, the effect on the children who would have stayed at home ($LATE_{hc}$ here) is an important, yet not identified and understudied, parameter of the preschool literature. We show that with a set of very plausible assumptions, we can derive bounds for the $LATE_{hc}$ and show that, after about 10 months of preschool, a child performance increases between 0.13 SD and 0.45 SD. Using additional instrumental variables and under an arguably heavier assumption (constant treatment effect within compliers type), the $LATE_{hc}$ is positive and significant, and is likely to be around 0.2 SD. This result corresponds to a third of the cognitive gap measure at baseline.

Third, our results on inequality raise an important caveat for supply-side preschool interventions. We show that, when implemented without any additional demand-side interventions (information campaign, in-kind, or cash incentives to enroll), a preschool program is unlikely to reduce the initial cognitive gap between poorer and wealthier households. Indeed, relatively wealthier parents are more likely to enroll

their children in preschool, and consequently, they are the primary beneficiaries of such programs. Specifically, we identified a large group of never-taker families (44% of the sample) that never benefits from any preschool and are significantly poorer and less educated than the rest of the sample. Given that these never-takers are expected to show large return to preschool (since they are staying at home in the absence of the program), finding ways to attract these families should be a top priority of governments to both increase the performance of preschool and significantly reduce cognitive inequality.

Our results directly relate to the existence 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, Gambia, and Indonesia (Bouguen et al., 2018, 2013; Blimpo and Pugatch, 2017; Brinkman et al., 2017). While implementation issue 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 ITT 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 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 child care provision exists would prove to be a very effective education policy.

This article also relates to the relatively recent literature on close substitute programs and on the identification of subLATEs (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 context, is reliable and is based on very plausible assumptions. Yet, extracting bounds, as well as implementing alternative identification strategies, depends 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. In our case, while we designed our experiment with an expected CPS take-up scenario of about 36%, the take-up on children going to any preschool barely reached 11% and could have led to type II errors. Luckily, lower than expected intra-cluster correlation (ICC)[40] and very high covariates predictive powered[41] allow us to implement our bounding strategy with success. Yet, in other

[40]For the main five test scores, ICC varied between 2.6% and 5.7% without control variables while our power calculations were based on an expected ICC of 10%.

[41]For 4 out of 5 test scores, the adjusted R^2 is between 53% and 67% while we expected 30%.

context, predicting more carefully the substitution patterns ex-ante – through pilot studies or a careful analysis of the available substitution offers – is critical to precisely isolate the subLATE parameters.

We will expand this research in different directions: first, we intend to address the issue of cognitive inequality using two additional treatment branches. In addition to constructing preschools, we implemented two variations of a demand-side intervention, targeted to stimulate preschool participation of the poorest (Berkes et al., 2018a). In the future, we also intend to work with the Cambodian government to develop other strategies in that direction (CCT, nudges, free meals, and so on). Second, in this article, we only provide a general description of the differences between the different preschools. In another article, we intend to provide a much more precise description of the teaching content by analyzing the videos and class visits we collected in every preschool of our sample. Third, we intend to follow up on every child in our sample, one year after the end of preschool, and again three years after the end of preschool when all the sampled children will be enrolled in primary school. Given the disappointing results found by Ozler et al. (2018) one year after the end of the intervention, tracking children in the long run will be an important complementary analysis.

The R^2 for the socio-emotional development test is lower (22%), potentially explaining why this dimension is often less significantly different from zero.

References

Alkire, S. and Santos, M. E. (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., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (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., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal Child Care, Maternal Labor Supply, and Family Well-Being. *Journal of Political Economy*, 116(4):709–745.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india*. *The Quarterly Journal of Economics*, 122(3):1235–1264.

Berkes, J., Bouguen, A., and Filmer, D. (2017). *Increasing Early Childhood Care and Development Through Community Preschools in Cambodia: Evaluating the Impacts*. AEA RCT Registry, `https://www.socialscienceregistry.org/trials/1045/history/22506`.

Berkes, J., Bouguen, A., and Filmer, D. (2018a). Attracting the right compliers: Evidence from a preschool construction program in cambodia. *work-in-progress*.

Berkes, J., Raikes, A., Bouguen, A., and Filmer, D. (2018b). Joint effects of parenting and nutrition status on child development: Evidence from rural cambodia.

Blimpo, M. P. and Pugatch, T. (2017). Scaling up children'school readiness in the gambia: Lessons from an experimental study. *Working paper*.

Bouguen, A., Filmer, D., Macours, K., and Naudeau, S. (2013). Impact evaluation of three types of early childhood development interventions in cambodia (english). *Policy Research working paper*, IE 97(WPS 6540).

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

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

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., and Pradhan, M. (2017). The impact of expanding access to early childhood education services in rural indonesia. *Journal of Labor Economics*, 35(S1):S305–S335.

Burde, D. and Linden, L. L. (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics*, 5(3):27–40.

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

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. CReAM Discussion Paper Series 1808, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

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

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

Cunha, F., Heckman, J. J., and Schennach, S. M. (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.

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

Elango, S., García, J. L., Heckman, J. J., and Hojman, A. (2015). *Early Childhood Education*, pages 235–297. University of Chicago Press.

Feller, A., Grindal, T., Miratrix, L., and Page, L. C. (2016). 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., Prado, E., Kariger, P., Raikes, A., et al. (2017). A toolkit for measuring early childhood development in low and middle-income countries. *World Bank Publications*.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., and Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science*, 344(6187):998–1001.

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

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

Heckman, J., Hohmann, N., Smith, J., and Khoo, M. (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. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

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

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

Kazianga, H., Levy, D., Linden, L. L., and Sloan, M. (2013). The effects of" girl-friendly" schools: Evidence from the bright school construction program in burkina faso. *American Economic Journal: Applied Economics*, 5(3):41–62.

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

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Lall, A. and Sakellariou, C. (2010). Evolution of education premiums in cambodia: 1997–2007. *Asian Economic Journal*, 24(4):333–354.

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

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., and Downer, J. (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.

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

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

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

A 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. Schools are usually better equipped, and the teacher is better trained and paid than community teachers. Classes last 3 hours, 5 days a week.

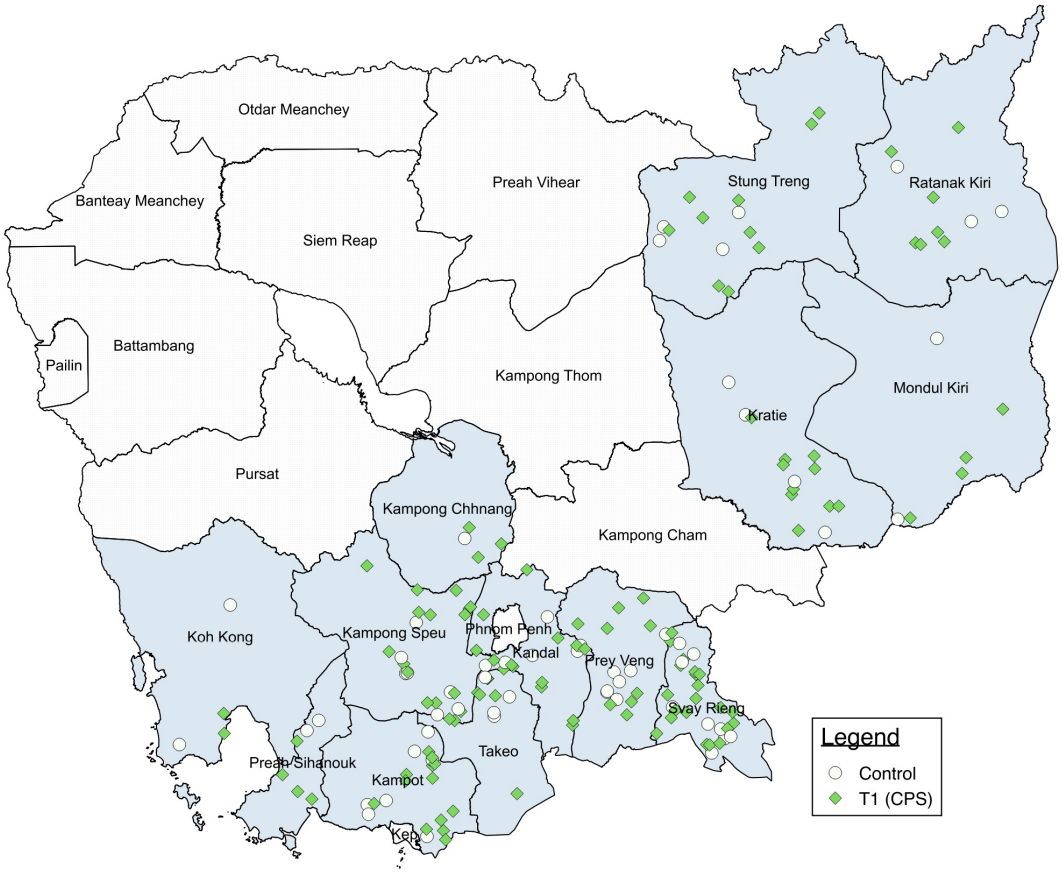
Figure 2: Informal Preschool (IPS)

Note: Informal preschool classes are usually given at the community teacher's house (under the house). CPS teachers receive a lower salary than SPS teachers. The class lasts 2 hours (usually in the morning).

Figure 3: Community Preschool (CPS)

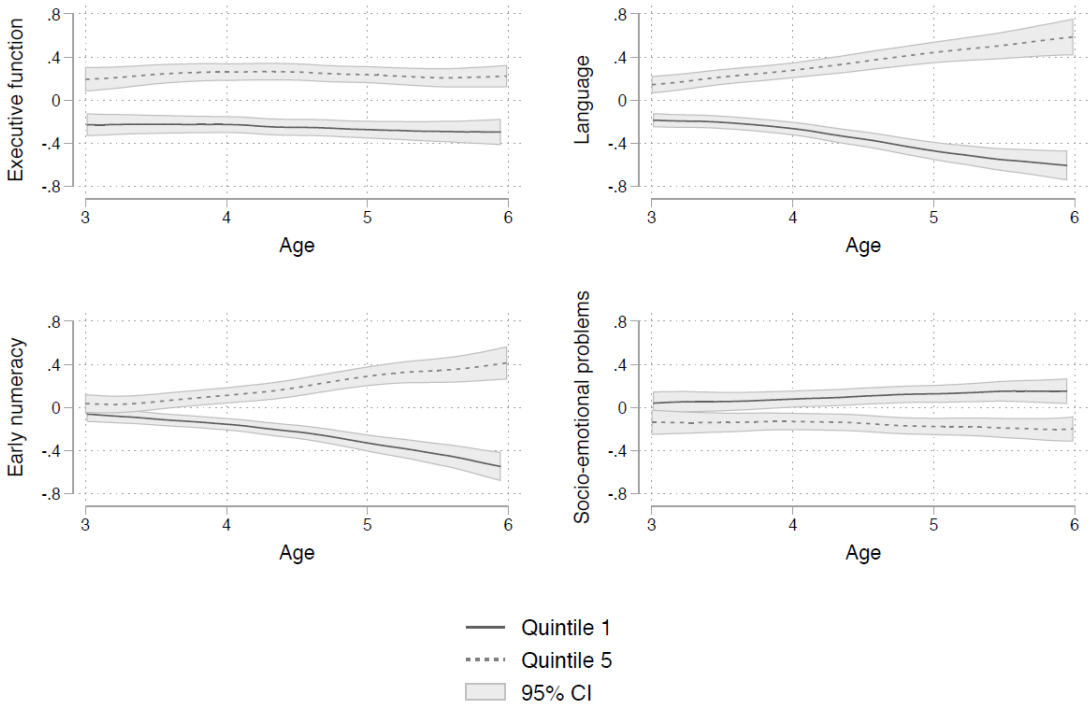
Note: Community preschools were constructed under the intervention studied in this paper. They were built under GPE II. All CPS buildings are the same, and they are better equipped than informal preschools. Newly recruited teachers receive better training and usually higher wages. Class lasts for 2 hours each day.

Figure 4: Location of Treatment and Control Villages in Cambodia



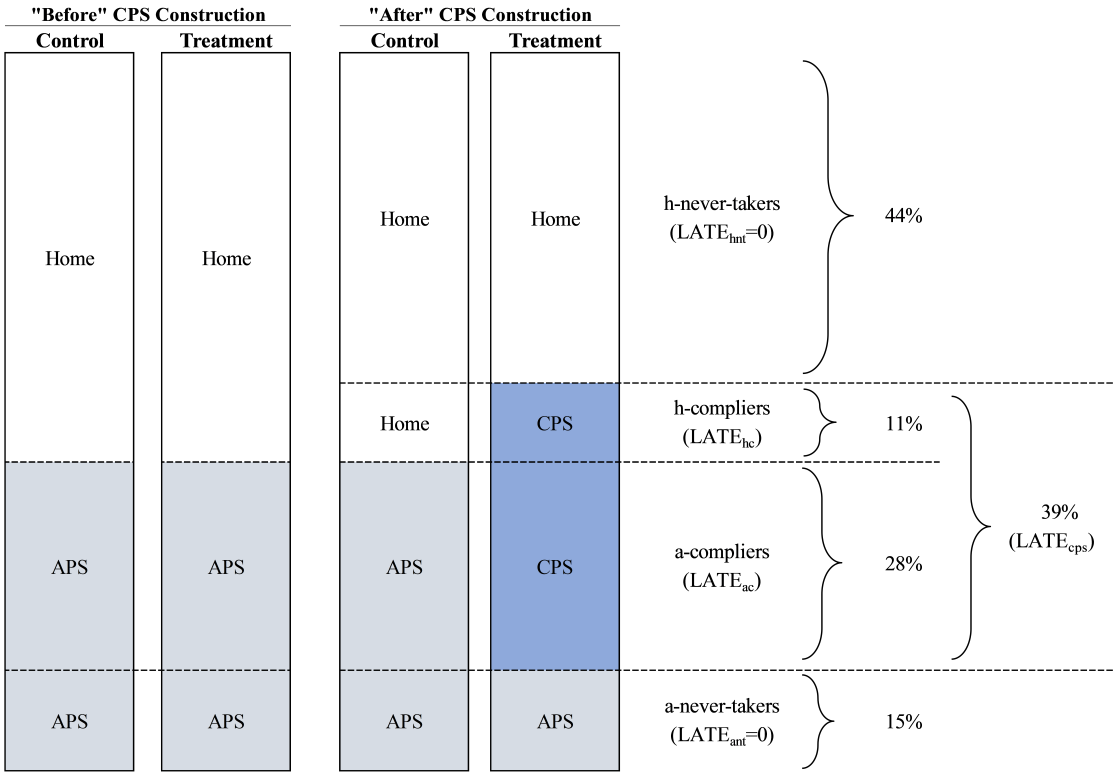
Note: Map shows sample villages by treatment status in the 13 provinces.

Figure 5: Baseline Cognitive Gap between Children from Wealthy and Poor backgrounds.



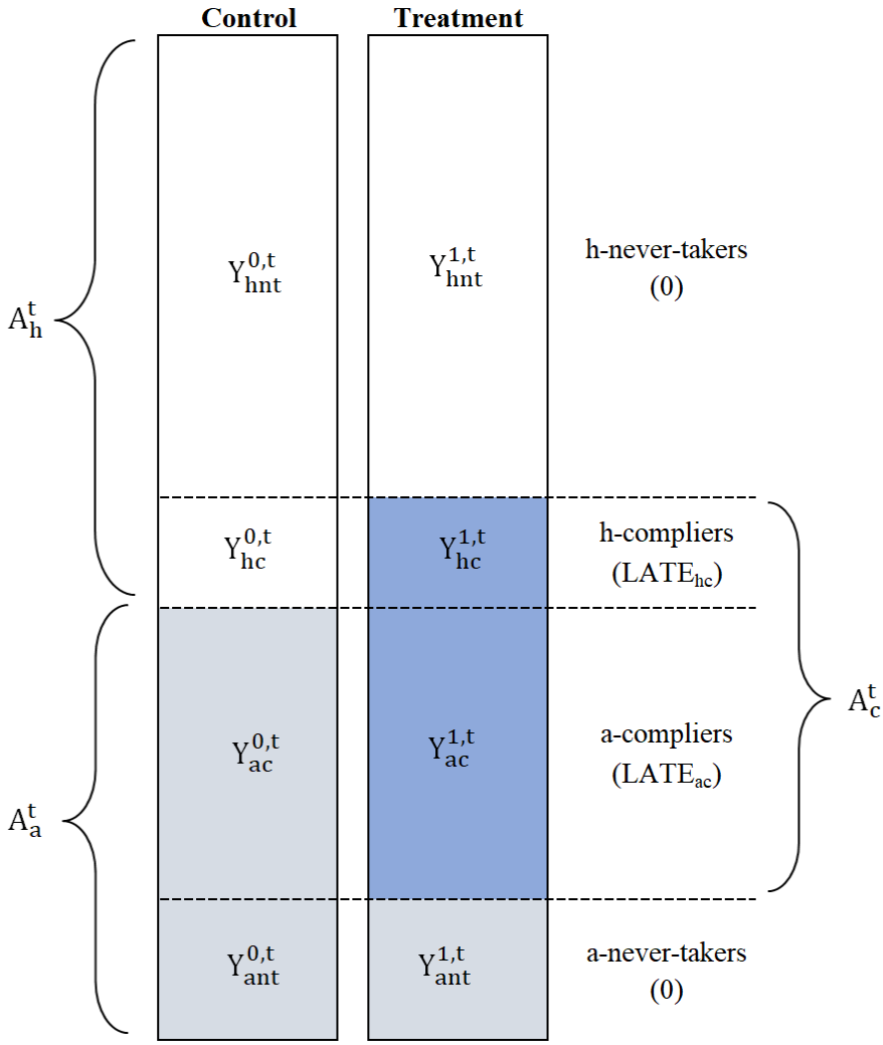
Note: Figure 5 shows gap in age-independent test scores between highest and lowest wealth score quintile. Age-independent test scores are standardized residuals from a regression of test score on age polynomials. See Berkes et al. (2018b) for details on the wealth measure. Note that in Figure 5 the socio-emotional problems is based on the Strength and Difficulty Questionnaire not on the socio-emotional score

Figure 6: Principal Strata Before and After CPS Construction



Note: Figure 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 7: Principal Strata Notations



Note: Figure shows the notation for each principal strata in treatment and control group after the construction of CPS ($Y_p^{z,t}$) as well as the notation for the weighted average ($A_d^{z,t}$)

B Tables

Table 1: Timetable

Period	**Activity**	**CPS construction**
03/2016	Begin CPS construction	0% completed
05/2016 - 07/2016	Baseline data collection	17% completed
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	0.0%	120	58	0.0%
Households	3757	8.7%	2578	1179	1.6%
Household members	20485	7.9%	14080	6405	1.2%
children from 3-5	4018	8.5%	2762	1256	3.8%
Tested children	3963	8.2%	2721	1242	4.5%
Baseline & midline					
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). Attrition and differential attrition column gives the respective overall attrition rate and the differential between treatment and control attrition. Endline data are expected to be available during the summer 2018.

Table 3: Treatment-Control Difference at Baseline – Household and Caregiver Data

	Baseline Sample			Baseline & Midline 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 4: 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 5: 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 6: 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

Table 7: Reasons for Non-Enrollment in Preschool, Follow-up

	Reason Applies	Most Important
Afraid child gets hurt on way to school	76%	17%
Child too active / not enough supervision	56%	7%
No one there to bring and pick up child	56%	17%
School is too far away	45%	6%
Child refuses to go / cries / is afraid	44%	11%
Child does not speak well enough	41%	6%
Too young (no detailed reason)	24%	17%
Enrollment was turned down	22%	13%
School facility is not adequate	11%	0%
School construction is not yet finished	10%	1%
Teacher is not present / cancels too often	9%	1%
Child has long-term illness / disability	9%	2%
Teacher is not well qualified	7%	0%
Child does not need any preschool	6%	0%
Did not think about sending child to preschool	5%	0%
Other	4%	3%
School is too expensive	2%	0%
Child must help with household chores	1%	0%
Personal disputes with teacher	1%	0%
		100%
Observations	3333	3333

Table 7 provides the reasons given by caregivers at follow-up to explain why their 3-5 year old children is not enrolled in any preschool (asked only in places where a preschool exists). We first asked whether each reason applies, then among the reasons that applies what is the most important one.

Table 8: Principal Stratum Names and Notations

		z=1		
		a	h	c
z=0	a	a-never-takers (ANT) $Y_{ant}^{z,t}$	a-defiers $\emptyset$	a-compliers (AC) $Y_{hc}^{z,t}$
	h	ha-defiers $\emptyset$	h-never-takers (HNT) $Y_{hnt}^{z,t}$	h-compliers (HC) $Y_{hc}^{z,t}$
	c	ca-defiers $\emptyset$	ch-defiers $\emptyset$	c-always-takers, $\emptyset$

Table 8 gives the name, acronym and the expected value notation of each principal strata i.e. for each value of Z and D.

Table 9: Non-parametric Descriptive Statistics of the Principal Strata, at Follow-up

	$Y_{hnt}^{0,1}$	$Y_{ant}^{0,1}$	$Y_{hnt}^{0,1} - Y_{ant}^{0,1}$	$Y_{hc}^{0,1}$	$Y_{ac}^{0,1}$	$Y_{hc}^{0,1} - Y_{ac}^{0,1}$
Caregiver variables						
Gender	0.907	0.922	-0.016	0.867	0.915	-0.048
	(0.009)	(0.013)	(0.015)	(0.082)	(0.018)	(0.081)
age	40.844	40.507	0.337	47.51	43.637	3.873
	(0.637)	(0.87)	(0.884)	(5.514)	(1.539)	(5.106)
Caregiver years of education	3.961	4.45	-0.489**	4.761	4.61	0.151
	(0.14)	(0.228)	(0.233)	(1.369)	(0.374)	(1.309)
Caregiver Raven Score	-0.066	0.146	-0.212***	0.216	-0.03	0.247
	(0.039)	(0.057)	(0.065)	(0.331)	(0.097)	(0.309)
Cognitive parenting	-0.064	0.151	-0.216***	-0.389	0.224	-0.613**
	(0.032)	(0.06)	(0.063)	(0.29)	(0.11)	(0.27)
Negative parenting	0.123	-0.009	0.132**	-0.225	-0.076	-0.149
	(0.037)	(0.05)	(0.057)	(0.335)	(0.079)	(0.319)
Socioemotional parenting	0.042	-0.046	0.087	-0.147	0.027	-0.174
	(0.038)	(0.052)	(0.055)	(0.329)	(0.08)	(0.321)
Children variables						
Executive functions	-0.401	0.813	-1.214***	-0.022	0.315	-0.337
	(0.033)	(0.042)	(0.057)	(0.292)	(0.077)	(0.297)
Language	-0.408	0.899	-1.306***	0.029	0.262	-0.232
	(0.028)	(0.083)	(0.089)	(0.247)	(0.119)	(0.266)
Numeracy	-0.476	0.875	-1.351***	0.058	0.344	-0.286
	(0.025)	(0.079)	(0.082)	(0.225)	(0.113)	(0.235)
Fine Motor	-0.382	0.955	-1.337***	-0.288	0.28	-0.568**
	(0.027)	(0.06)	(0.07)	(0.247)	(0.111)	(0.259)
Socioemotional skills	-0.186	0.308	-0.494***	-0.232	0.284	-0.516*
	(0.036)	(0.06)	(0.063)	(0.338)	(0.095)	(0.313)

Table 9 gives the estimation for $y_{hnt}^{0,1}$, $y_{ant}^{0,1}$ $y_{hc}^{0,1}$ and $y_{ac}^{0,1}$ at follow-up. $y_{hnt}^{0,1}$ and $y_{ant}^{0,1}$ are directly observed. Their comparison is given in column $y_{hnt}^{0,1} - y_{ant}^{0,1}$. $y_{hc}^{0,1}$ and $y_{ac}^{0,1}$ are non-parametrically estimated using equation 5. Their standard errors are calculated using the delta method. Column $y_{hc}^{0,1} - y_{ac}^{0,1}$ gives the comparison between h and a compliers in the control group at follow-up.
* 10%, ** 5%, *** 1% significance level

Table 10: Village Infrastructure at Follow-up

	Obs.	C	T-C
Any preschool in village	178	0.81	0.123**
			(0.057)
Any community preschool in village (CPS)	178	0	0.858***
			(0.032)
Any alternative preschool in village (APS)	178	0.81	-0.552***
			(0.066)
$\cdots$ Informal preschool in village (IPS)	178	0.655	-0.564***
			(0.068)
$\cdots$ State preschool in village (SPS)	178	0.241	-0.058
			(0.067)
$\cdots$ SPS class size in villages with SPS†	36	6.429	-0.974
			(1.558)

Table 10 presents village level regressions of the outcome variable in line against the treatment variable. Estimates correct for heteroskedasticity. 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.

† : number of sampled children enrolled in SPS in villages with a SPS

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

Table 11: Enrollment and ITT Exposure

	Full sample			5-Year-olds		
	Obs.	C	T-C	Obs.	C	T-C
Enrollment rate						
Any school	4011	0.435	0.106***	1153	0.683	0.099**
			(0.036)			(0.043)
CPS	4011	0	0.389***	1153	0	0.473***
			(0.025)			(0.032)
Alternative preschool, APS	4011	0.435	-0.283***	1153	0.683	-0.374***
			(0.034)			(0.045)
$\cdots$ Informal preschool (IPS)	4011	0.284	-0.247***	1153	0.397	-0.32***
			(0.034)			(0.047)
$\cdots$ State preschool (SPS)	4011	0.112	-0.041*	1153	0.18	-0.052
			(0.024)			(0.04)
$\cdots$ Primary school	4011	0.04	0.004	1153	0.106	-0.002
			(0.009)			(0.022)
Months of Exposure						
Any school	4006	3.672	1.01**	1149	6.94	1.289*
			(0.401)			(0.69)
CPS	4006	0	3.421***	1149	0	5.205***
			(0.253)			(0.406)
Alternative preschool, APS	4006	3.672	-2.411***	1149	6.94	-3.916***
			(0.347)			(0.646)
$\cdots$ Informal preschool, IPS	4006	2.411	-2.088***	1149	4.193	-3.402***
			(0.32)			(0.578)
$\cdots$ State preschool, SPS	4006	0.994	-0.426*	1149	1.968	-0.744*
			(0.219)			(0.45)
$\cdots$ Primary school	4006	0.267	0.104	1149	0.779	0.23
			(0.074)			(0.24)

Table 11 gives the ITT first stage fore several measures of preschool participation and for the full sample and the sample of 5-years-olds. 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. 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 12: ITT Impacts on Children Performance.

	Full sample				5 Year-olds			
	Obs.	C	(1)	(2)	Obs.	C	(1)	(2)
Cognitive Development (CD)								
Executive functions	3963	0	0.042	0.05*	1138	0.738	0.101*	0.074*
			(0.044)	(0.026)			(0.056)	(0.044)
Language	3963	0	0.027	0.046	1138	0.783	0.1	0.077
			(0.049)	(0.03)			(0.087)	(0.062)
Numeracy	3963	0	0.029	0.049*	1138	0.777	0.159*	0.169***
			(0.048)	(0.029)			(0.091)	(0.064)
Fine motor	3963	0	0.071	0.061**	1138	0.832	0.142*	0.072
			(0.046)	(0.03)			(0.076)	(0.056)
Socio-emotional	3959	0	0.024	0.05	1138	0.369	0.084	0.109**
			(0.057)	(0.037)			(0.067)	(0.054)
CD Index	3963	0	0.039	0.051***	1138	0.7	0.117**	0.1***
			(0.04)	(0.02)			(0.059)	(0.038)
Physical Development (PD)								
Height for age	3934	-1.703	0	-0.003	1111	-1.724	0.039	0.024
			(0.044)	(0.035)			(0.072)	(0.062)
Weight for age	3934	-1.533	0.043	0.056*	1111	-1.65	0.056	0.08
			(0.037)	(0.033)			(0.069)	(0.066)
Gross Motor	3963	0	-0.008	0.006	1138	0.702	-0.015	-0.036
			(0.058)	(0.03)			(0.067)	(0.05)
PD index	3969	-1.068	0.009	0.019	1140	-0.847	0.014	0.023
			(0.028)	(0.026)			(0.052)	(0.046)

Table 12 gives the ITT treatment effect estimates for each outcome variable. Estimates correct for heteroskedasticity and are clustered at the village level. Column *Obs.* indicates the number of children who took the test at follow-up. Column *C* the performance of the children in the control villages, standardized using the control group (except for anthropometrics measures). Column *(1)* gives the results without covariates, *(2)* with covariates.
* 10%, ** 5%, *** 1% significance level

Table 13: ITT impact on Parenting and Perceived Return to School

	Full sample				5 Years old			
	Obs.	C	(1)	(2)	Obs.	C	(1)	(2)
Parenting Score (PS)								
Negative parenting	4012	178	0.026	0.016	1138	0.702	0.021	0.015
				(0.037)			(0.073)	(0.059)
Socio-emotional parenting	4012	178	0.006	0.018	1140	-0.847	0.023	0.034
				(0.044)			(0.071)	(0.066)
Cognitive parenting	4012	178	0.048	0.067	1153	-0.058	0.11	0.109*
				(0.041)			(0.079)	(0.063)
PS index	4012	178	0.009	0.023	1153	-0.174	0.037	0.043
				(0.026)			(0.049)	(0.038)
Parental Perception								
Optimal preschool age	4010	178	4.03	-0.134**	1153	-0.042	-0.118	-0.115
				(0.053)			(0.082)	(0.072)
Optimal primary school age	4010	178	5.912	-0.045*	1153	-0.053	-0.053	-0.045
				(0.026)			(0.044)	(0.043)
Perceived Income no school	4010	178	104.45	1.278	1152	4.294	-1.991	1.1
				(2.642)			(5.112)	(4.175)
Perceived Income full prim. school	4010	178	148.227	6.455*	1152	6.066	1.277	2.903
				(3.443)			(9.32)	(9.266)
Perceived Income full second. school	4010	178	240.786	16.585***	1152	103.385	17.173	19.999
				(5.925)			(16.146)	(15.251)

Table 13 gives the ITT treatment effect estimates for each outcome variable. Estimates correct for heteroskedasticity and are clustered at the village level. Column *Obs.* indicates the number of children who took the test at follow-up. Column *C* the performance of the children in the control villages, standardized using the average in the control group (except for anthropometrics measures). Column *(1)* gives the result of the regression without any control, *(2)* with the complete set of control variables.
* 10%, ** 5%, *** 1% significance level

Table 14: $LATE_{hc}$ Bounds at Follow-up

	Full Sample			5 year-olds		
	Obs.	Low Bound	High Bound	Obs.	Low Bound	High Bound
Any School Exposure (m)	4006	2.71***	9.537***	1149	2.758**	13.101***
		(0.815)	(1.672)		(1.176)	(3.771)
Cognitive Development (CD)						
Executive functions	3959	0.126*	0.439*	1138	0.152*	0.727*
		(0.065)	(0.224)		(0.091)	(0.432)
Language	3959	0.117	0.408	1138	0.16	0.766
		(0.075)	(0.267)		(0.128)	(0.604)
Numeracy	3959	0.127*	0.441	1138	0.349***	1.667**
		(0.072)	(0.271)		(0.131)	(0.769)
Fine motor	3959	0.152**	0.527*	1138	0.15	0.716
		(0.075)	(0.271)		(0.114)	(0.54)
Socio-emotional	3959	0.127	0.442	1138	0.207***	1.077
		(0.092)	(0.342)		(0.079)	(0.675)
CD index	3959	0.13**	0.451**	1138	0.275***	0.991**
		(0.051)	(0.195)		(0.101)	(0.441)
Parenting Score (PS)						
Negative parenting	4011	0.036	0.126	1153	0.031	0.15
		(0.092)	(0.325)		(0.12)	(0.567)
Socioemotional parenting	4011	0.037	0.13	1153	0.07	0.335
		(0.109)	(0.38)		(0.136)	(0.641)
Cognitive parenting	4011	0.168	0.592	1153	0.227*	1.087*
		(0.103)	(0.367)		(0.13)	(0.654)
PS Index	4011	0.056	0.199	1153	0.089	0.424
		(0.065)	(0.226)		(0.077)	(0.381)

Table 14 gives the bounds for the $LATE_{hc}$. The low bound is the $LATE_{cps}$, the high bound is the $LATE_{ps}$. We provide the bounds with a full set of control variables. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Table 15: Bounds and Conditional LATE

	Z	Z	Z x prov. FE & vill. infra.			Z x prov. FE, & vill. infra. & Individual covariates		
	$LATE_{cps}$ (low bound)	$LATE_{ps}$ (high bound)	$LATE_{cps}$	$LATE_{hc}$	$LATE_{ac}$	$LATE_{cps}$	$LATE_{hc}$	$LATE_{ac}$
Executive function	0.126*	0.439*	0.091	0.087	0.093	0.096*	0.196*	0.050
	(0.065)	(0.224)	(0.056)	(0.123)	(0.075)	(0.053)	(0.112)	(0.069)
Overid. test p-value			0.190	0.147		0.045	0.040	
Language	0.117	0.408	0.093	0.136	0.072	0.129*	0.218*	0.087
	(0.075)	(0.267)	(0.067)	(0.126)	(0.092)	(0.066)	(0.116)	(0.089)
Overid. test p-value			0.894	0.859		0.706	0.785	
Numeracy	0.127*	0.441	0.088	0.124	0.072	0.132*	0.239**	0.082
	(0.072)	(0.271)	(0.067)	(0.123)	(0.090)	(0.070)	(0.111)	(0.090)
Overid. test p-value			0.221	0.176		0.199	0.208	
Fine motor	0.152**	0.527*	0.131*	0.029	0.179**	0.139*	0.055	0.179**
	(0.075)	(0.271)	(0.073)	(0.163)	(0.085)	(0.074)	(0.148)	(0.087)
Overid. test p-value			0.709	0.823		0.797	0.810	
Socio-emotional	0.127	0.442	0.114	-0.008	0.172	0.146*	0.084	0.175
	(0.092)	(0.342)	(0.078)	(0.170)	(0.116)	(0.079)	(0.154)	(0.114)
Overid. test p-value			0.512	0.480		0.413	0.386	
Children index (SUR)	0.163**	0.568**	0.130**	0.093	0.148*	0.162***	0.199*	0.144**
	(0.064)	(0.245)	(0.058)	(0.127)	(0.076)	(0.057)	(0.112)	(0.072)
Overid. test p-value			0.372	0.324		0.249	0.216	
First-stage F	238.684	11.161	48.922	12.237	9.145	35.737	11.509	6.005

Table 15 gives for each child outcome, the bounds from table 14 in the first two columns, then gives the estimates of $LATE_{cps}$, $LATE_{hc}$ and $LATE_{ac}$ based on the approach described in Section 3.5. In the first panel, X includes Z as well as its interaction with X (province fixed effect, village population size, village land area, presence of a primary and secondary school) in the second panel, X includes in addition, household, caregiver and child baseline variables (education level, raven test score, poverty dummy and all individual baseline test score). For each LATE(X) we provide the p-value of the over-identification test and the Angrist and Pischke (2009) F statistics of the first stage regressions. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Table 16: Heterogeneous Preschool Impact

	Baseline Wealth Score		
	Q1	Q5	Q5-Q1
Enrollment			
CPS	0.304***	0.39***	0.087*
	(0.034)	(0.038)	(0.215)
APS	-0.263***	-0.297***	-0.034
	(0.05)	(0.051)	(0.26)
Children performance			
Executive functions	0.033	0.12**	0.086
	(0.059)	(0.051)	(0.079)
Language	0.011	0.074	0.063
	(0.044)	(0.068)	(0.08)
Early Numeracy	-0.018	0.135*	0.153*
	(0.052)	(0.072)	(0.091)
Fine motor	0.046	0.057	0.011
	(0.051)	(0.06)	(0.072)
Socio-emotional	0.128	0.005	-0.123
	(0.092)	(0.076)	(0.111)
Children index	0.05	0.099*	0.048
	(0.049)	(0.053)	(0.07)

Table 16 gives the ITT impact on children outcome among the first and last quintile of the baseline wealth index. Column Q5-Q1 gives the comparison between Q1 and Q5 using a Seemingly Unrelated Regression model that estimate conjointly Q1 and Q5 regressions. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

C Appendix: Validity of A4 (Exclusion Restriction)

As touched upon in section 3.2, assumption A4, and especially sub-assumption A4 (ii) and (iii), are subject to violations if the construction of a CPS affected 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.

First, in term of class-size, we have reasons to believe that this problem is unlikely. Indeed, as seen in Table 10, the construction of a CPS generally entails the IPS to shut 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: IPS are therefore unlikely to have been indirectly benefited from a class size reduction.[42] Yet, since SPS were not shut down when a CPS was constructed, SPS are more likely to have been indirectly affected by the CPS construction. We look at this possibility in the last row of Table 10. Since we have not collected the class size in SPS, we use as a proxy the average number of sampled children enrolled in SPS class per village. Since we did not sample all children in the village (but in average 26 children), we divide this number by our average sample weight (here estimated at 53%). The last row of Table 10 indicates that, on average, the number of sampled children enrolled in SPS in the 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 $1.83 * 0.083 * 0.03 = 0.4pp$ maximum. This would correspond to about 7.8% of the overall treatment effect (5.1% of a SD). Reduction of the class size in SPS is hence unlikely to significantly modify the magnitude of the treatment effect.

Second, peer composition may violate the exclusion restriction if, for instance, CPS attract specific children, leaving SPS or the remaining IPS with more homogeneous or better/worst peers. As mentioned in the body of the text, since APS are composed of better quality school, we would expect the peer composition to have improve in APS and as a result the a-never-taker to benefit from more favorable conditions. We test this possibility in Table C.1 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 suggest that CPS constructions have not modified a-never-takers' composition.[43]

Lastly, the exclusion restriction may be violated if the CPS construction modified the involvement of never-taker parents (h and/or a). We already know that the program positively impacts parents' perceptions and self-reported parenting practices, see Table 13. Yet, such effect does not constitute a A5 violation as long as it only affects the a- and h-compliers.[44]. A5 would be violated however if the parenting

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

[43]Again, the same cannot be done for IPS, as the CPS construction forced many IPS to shut down (see Table 10), and therefore, the IPS treatment sample is a selected one.

[44]It should be reminded that our experiment measures the overall effect of a preschool construction, including indirect effects on parental perception and involvement

effect expands to h-and a-never-takers.

We look at this possibility in Table C.2 where we estimate the ITT effect on SPS never-takers children. Table C.2 shows that parents do not report different perceived returns to education, that children do not perform better but that parents do declare being in average more involved 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. This would shed doubts on the results presented in Table 13 but would not fundamentally affects the validity of the experiment. Yet, as suggested in the body of the text, it could also be that all parents changed their behavior towards early education because of the construction. If that was 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., 2018a), 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. (2018a), children's performance increases by a maximum of 10 pp for each standard deviation increase of the cognitive parenting index. A 11 pp effect would hence 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 C.2 is evidence that Assumption A5 is valid as a whole. Indeed, class-size, peer composition and parental involvement are all forces that would bias upward the impact in Table C.2 in case of violation of A5. With an overall ITT effect on SPS children estimated at -0.02 SD, we are confident that our experiment is not concerned by a violation of the exclusion restriction.

Table C.1: 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 C.1 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

Table C.2: 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 C.2 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