Formal kindergarten improves cognitive but not socio-emotional development in India*

Joshua Dean
University of Chicago Booth School of Business

Seema Jayachandran
Northwestern University

September 3, 2019

Abstract

In many developing countries, formal kindergarten is available but has low participation, which exacerbates the socioeconomic gradient in readiness for primary school. We study the effects of expanding access to kindergarten on child development in Karnataka, India, through a randomized evaluation. We partnered with a private kindergarten provider to offer two-year scholarships to low-income children. Children who attend our partner's kindergarten due to the scholarship experience a 0.8 standard deviation gain in cognitive development. This effect size is considerably larger than most estimates for Head Start in the US. About 40% of the effect persists through the end of first grade. We find no effect on socio-emotional development. The scholarship causes children to attend our partner's kindergarten instead of attending (a) no kindergarten or (b) a different kindergarten. We use machine learning techniques to predict each child's counterfactual activity and then estimate separate treatment effects for different types of switchers. The short-run effect on cognition is driven by children who would not have otherwise attended kindergarten, but the persistent component of the effect does not vary by counterfactual.

Keywords: early childhood education, **JEL Codes:** I24, I25, J13, O12.

*We thank UBS Optimus Foundation for funding this research; Tvisha Nevatia, Sadish Dhakal, Aditya Madhusudan, Akhila Kovvuri, Sachet Bangia, Alejandro Favela, Ricardo Dahis, and Ariella Park for research assistance; and Azzurra Ruggeri for advising us on measurement of child development. This study was preregistered in the AEA Trial Registry as AEARCTR-0001078.

1 Introduction

With many developing countries making significant progress toward universal enrollment in primary education in recent decades, expanding access to formal pre-primary education has emerged as one of the next major policy priorities. Government-provided or subsidized kindergarten and pre-kindergarten remain rare in developing countries. At the same time, there is a growing consensus on the importance of this pre-primary period for children's development (Almond and Currie, 2011).

With limited budgets, governments are interested in whether expanding access to formal kindergartens, with highly structured curricula delivered by teaching professionals, is a cost-effective way to improve learning outcomes. However, most evidence on the impacts of formal pre-primary school is from developed countries, but there are several reasons to be cautious about extrapolating this evidence to developing countries. First, the quality of instruction might differ. Second, the effect depends on the counterfactual, and these are quite different across contexts; for example, the level of stimulation children receive at home might differ, and whether they are cared for at home or in day care centers will vary across contexts. Third, the persistence of gains from pre-primary school will depend on the emphasis that primary schools put on helping the lagging students catch up; the fact that in developing countries, there is often less emphasis on remedial instruction and more "teaching to the top" means that effects might be particularly persistent in this context.

In this paper we partner with a large, private provider of kindergarten, Hippocampus Learning Centers (HLC), in Karnataka, India to evaluate how attending formal kindergarten affects cognitive and socioemotional development. HLC runs kindergartens in over 200 villages. We randomize scholarships to attend HLC at 10-20% of the typical cost among a sample of 808 children spread across 71 villages in 10 districts of Karnataka where HLC operates. The scholarship offer increases the probability of attending a kindergarten by 20 percentage points and the probability of attending HLC by 47 percentage points.

We find that attending HLC has a large positive effect on cognitive development in the short run. The effect size attenuates by the end of first grade, but the persistent component is still sizeable. However, there is no impact on socio-emotional development. More specifically, at the end of kindergarten, children induced into attending HLC by the scholarship offer score 0.8 standard deviations higher than their peers on cognitive tests. This is the same magnitude as the growth in scores among those in the control group who do not attend any

kindergarten. These benefits are broadly distributed across all test domains. Around 40% of the effect persists to the end of the first grade on tests common to both endlines. This is similar to the Head Start Impact study where approximately 45% of the effect persists from the end of the second year of Head Start through the end of first grade (Kline and Walters, 2016); although the Head Start decline starts from a much smaller effect size of 0.245.

We then decompose these effects by children's propensity to have attended some form of kindergarten in absence of our intervention. Understanding this heterogeneity is policy relevant for two reasons. First, we were logistically limited in our ability to target the intervention. Knowing the gains to those unlikely to have attended without the scholarship provides an upper bound to the effect that the government might see if they were able to roll out a similar voucher program with improved targeting. Second, we were also logistically limited to partnering with one provider, but in a scale-up policy makers would likely allow the vouchers to be used at multiple providers. The treatment effect on those who we induce to transfer to HLC from other providers is a measure of how well our provider outperforms the other options in the market. There are still remaining questions to be answered about the effects of a scaled program, particularly through possible expansion of the set of providers induced by increased demand, but this decomposition provides important guidence.

To accomplish this, we follow Crepon et al. (2019) in using machine learning on the control group to form counterfactual attendance predictions for each member of the sample based on demographic and village characteristics. Then we combine this approach with the method proposed by Hull (2018) to obtain LATEs for two exclusive types of compliers – those who would have attended some kindergarten and those that would have attended none. We find that while attending our provider benefits all students, the effects are primarily concentrated among those who would not have otherwise gone to kindergarten.

2 Background

2.1 Indian Formal Pre-Primary Education

Formal pre-primary education in India is quite limited, with a gross enrollment ratio of only 14% in 2017 (UNESCO Institute of Statistics). There is a country-wide network of Anganwadi centers, administered by the Ministry of Women and Child Development, but they provide primarily supervision and nutritional services. However, this might change soon as the central government announced in the summer of 2019 a draft proposal to include the

three years before pre-primary under the right to education act (Jebaraj, 2019). This policy as currently drafted would lead to a formalized pre-primary experience administered through public primary schools. In the meantime some individual states are already acting to expand their formal kindergarten capacity. For example, Karnataka announced in the summer of 2019 that it would be opening 276 formal pre-primary centers in existing primary schools (Belur, 2019).

2.2 Existing Evidence on Formal Pre-Primary Education

There is a large literature in the United States and other high-income countries assessing the benefits of formal pre-primary education (see Almond and Currie (2011) for an overview). In general, non-experimental methods in this literature have found quite large and persistent benefits to early intervention, while randomized evaluations have tended to find smaller effects. Although, there are some notable exceptions such as Heckman et al. (2010).

However, it is unclear to what extent we should be confident in extrapolating from these settings to low and middle income countries for two reasons. First, as Kline and Walters (2016) demonstrate, these effects are highly dependent on the counterfactuals faced by students who do not enroll. There are very good reasons to believe that these counterfactuals differ significantly across contexts and thus the treatment effects of expanding enrollment are also likely to differ. For example, compared to a child in the US, a child in rural India is much less likely to have a home with formal educational materials absent enrollment in pre-primary education, but is potentially more likely to have increased social interaction facilitated by Anganwadi centers. Second, many programs in developing countries suffer from significant implementation challenges. Finally, the degree and type of subsequent educational investments are likely to be quite different. For example, with No Child Left Behind the US government has had a significant stated priority on improving the outcomes of struggling students in primary school. In contrast, Indian schools often struggle to provide these remedial services. For example, Muralidharan et al. (2019) find no improvement over the course of a year on either Math or Hindi for students in the bottom tercile of middle-schools in Delhi.

In low and middle income countries there is significant work on the importance of the quality of early life experiences. For example, Gertler et al. (2014) and Attanasio et al. (2015) study the importance of appropriate cognitive stimulation in young children. Other work considers the importance of cash transfers during this period (e.g. Adhvaryu et al.

(2018)). This literature generally finds very large returns to investment or avoiding negative experiences.

In terms of kindergarten in developing countries, there are generally two sets of evidence. First, a large non-experimental literature examining a variety of programs of different types that finds mixed evidence (see Engle et al. (2011) for an overview). Second, experimental studies on community based preschools (e.g. Blimpo and Pugatch (2017), Martinez et al. (2017), Bouguen et al. (2018), and Berkes and Bougen (2019)). These schools are generally run for a few hours a day by a member of the community. In contrast, our provider runs schools much more similar to a U.S. style, formal preschool. They typically meet from 10:00 am to 4:00 pm, and there are daily, detailed schedules and lesson plans developed by curriculum specialists in the central office who stay abreast of current educational practice.[1] In order to accomplish this level of formality, teachers go through 20 days of intense training after being hired. They then meet with HLC curriculum staff once a month at regional gatherings to go over the lesson plans they will teach in the coming month. Finally, for each skill domain there are clearly specified learning targets and children are tested every month in order to assess their progress against these targets.

3 Experimental Design

3.1 Overview

We randomize scholarships to attend two years of kindergarten at HLC. HLC is a private kindergarten provider operating over 200 centers. Fees vary by village size and year of kindergarten from 3,500 to 4,800 rupees per year, and the total costs inclusive of materials vary from 5,125 to 7,825 rupees per year. The scholarship covers all fees and materials except for a 1000 rupee co-pay per year, which the family is required to pay. Thus the scholarship represents a 4,125 to 6,825 rupee subsidy (60 to 100 US dollars), which is 80% to 87% of total costs.

Families who do not enroll at HLC may choose one of the following alternatives: government run anganwadi centers, other private kindergarten providers, or to care for their children at home. The anganwadi centers are primarily focused on providing supervision and do not have a structured curriculum. Other private kindergartens have similar fees and

[1]For concreteness, in an example lesson plan from the teacher training materials, the teacher is instructed to use that day's time allocated to English to have children complete the letter tracing patterns on page 60 of the HLC writing book.

teacher salaries, and also have structured curriculums. HLC differs by offering smaller class sizes with bare-bones facilities and putting less emphasis on rote learning and homework.[2] They also target children from poorer families.

3.2 Enrollment and Randomization

In the spring of 2016 we enrolled 808 children across 71 villages in Karnataka in our study (Figure 1 shows a map of the study villages.). Within each village enumerators visited households to identify those with children between the ages of 3.5 and 4.5 on June 1, 2016 and who were not currently enrolled at HLC.[3] Surveyors then administered an asset inventory to assess households' economic standing. We then use a predetermined formula based on piloting to combine the assets into a score and use a predetermined cutoff to determine eligibility. For eligible households, surveyors then scheduled visits to complete a baseline assessment of the children's development and to survey the parents.

888 children met these inclusion criteria and 808 chose to enroll in the study by completing the baseline assessments and surveys. Within each village we enrolled no more than 16 children in the study to avoid overwhelming the HLC center and to minimize potential spillover effects. In cases where more than 16 children completed all enrollment criteria, we randomly selected 16 subjects. We then randomly assign half of the children in each village to receive a scholarship (see Table 2 for sample characteristics and balance tests).

3.3 Measuring Cognitive and Socioemotional Development

We assess the children's development at three time points: before kindergarten (baseline), after two years of kindergarten (endline 1), and after the end of first grade (endline 2). All tests were conducted at home in order to keep the testing environment consistent across children. We experience minimal attrition locating 796 children at the first endline and 786 at the second. Table 3 shows that these minimal attrition levels are uncorrelated with treatment status.

We use a variety of tests drawn from the fourth editions of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and the Developmental Indicators for the Assessment

[2]Our data on alternative private kindergartens comes from a survey that we ran of private kindergartens operating in our study villages.

[3]Our study considers the two standard years of kindergarten, but HLC also offers a "pre-kindergarten" year that some children may have already been enrolled in. Additionally, we intentionally recruit towards the end of the enrollment period so that families who are likely to send their children to HLC without the scholarship have already had time to enroll.

of Learning (DIAL), and individual sourced additional tests to assess children's reasoning, memory, languange, mathematics, creativity and motor skills. These tests are designed to be fun and engaging to minimize testing anxiety while getting at core skills. Figure 2 shows an example of a problem designed to assess children's reasoning abilities. Unfortunately, at the second endline, piloting revealed that the DIAL tests were no longer sufficiently difficult to provide meaningful variation. To compensate, we removed the DIAL instruments and added two new math assessments to maintain coverage of most domains.

Directly assessing children's socioemotional development at these ages is difficult without highly trained evaluators. At the first endline we rely on parents' reports using the standard "Strengths and Difficulties Questionnaire". This questionnaire asks parents to rate whether several statements about their child (such as "Restless, overactive, cannot stay still for long") are "not true", "somewhat true" or "certainly true". At the second endline when children are older we augment this with direct assessments of their motivation to learn, personalities, ability to read others emotions, prosociality and behavioral performance in school. Table 1 summarizes all measures administered to the children by domain, source and notes when they were employed.

4 Analysis

4.1 Main Results

We pre-specified two primary specifications, a reduced-form model and an instrumental variables one. The reduced-form model for the effect of the scholarship offer for each endline is as follows:

$$y_{ij} = \beta_1 \text{scholarship}_{ij} + \beta_2 \text{baseline}_{ij} + \text{village}_j + f(\text{gender}_{ij}, \text{age}_{ij}) + \epsilon_{ij} \tag{1}$$

where y_{ij} is an outcome for child i in village j, scholarship_{ij} is whether the child was offered a scholarship, baseline_{ij} is the baseline value of y_{ij} if available, village_j are fixed effects for the villages, and $f(\text{gender}_{ij}, \text{age}_{ij})$ are age and gender controls. Specifically, we pre-specified that if neither age and gender fixed effects nor an age cubic-polynomial interacted with gender improved explanatory power in the control group at endline, that we would use a non-interacted cubic polynomial in age with an indicator for gender for this set of controls. Based on this algorithm, we use the non-interacted controls throughout.

Second, we estimate the following IV, which pools across different counterfactual op-

tions:

$$y_{ij} = \beta_1 \text{enrolled in HLC}_{ij} + \beta_2 \text{baseline}_{ij} + \text{village}_j + f(\text{gender}_{ij}, \text{age}_{ij}) + \epsilon_{ij} \quad (2)$$

where being enrolled in HLC is now instrumented with the scholarship offer.[4]

Our outcomes are z-score indicies. For each test, we subtract off the control mean and divide by the standard-deviation.[5] We then average across all cognitive tests to form our primary outcome, within the domains shown in Table 1 to form domain-specific indices, and across the socio-emotional measures.

4.2 Heterogeneity by Propensity to Enroll

We have multiple types of compliers depending on what the child would do in absence of the scholarship. We can then think of the LATE estimated by the pooled IV above as a weighted average of these subLATEs. In this section we attempt to separately identify these counterfactual specific treatment effects.

Understanding how these effects differ is not only intellectually interesting, but also highly policy relevant. Consider a government contemplating expanding a voucher program similar to the scholarships that we randomize. The effect of our scholarship on children who otherwise wouldn't have attended kindergarten gives insight into how large of an effect the government might expect if they could perfectly target the program. The impact on switchers from other kindergartens then is also informative about the degree to which our provider is better than the average option in the market. This relative position is important for deciding how restricted the vouchers should be across providers.

We focus on a specific element of the fallback heterogeneity – the child's propensity to enroll in a kindergarten (either HLC or another private provider) absent the scholarship. Thus, we pool together always takers and KG compliers and we pool together children who would have stayed home with those that would have attended anganwadi centers.[6] We believe

[4]We deviate here from our pre-analysis plan, which stated we would use "enrolled in any kindergarten" as the endogenous variable. However, the scholarship also changes which kindergarten a child attends, constituting an exclusion restriction violation. We realized this is not an innocuous problem when we saw that a large share of children induced to attend HLC switch from other kindergarten providers.

[5]Note that before the first endline, we specified we would use the baseline, control mean and standard deviation. We realized that this makes it more difficult to interpret the results as test composition and the standard deviations change. Therefore before endline 2 we pre-specify that we would use contemporaneous control values.

[6]Ideally, we would like to be able to separately identify these effects. Unfortunately, we have very few children who either stay home or go to HLC absent the scholarship making it impossible to separate them.

this captures the most important aspect of the heterogeneity, whether a child is receiving a formal school curriculum or not.

One approach to measuring counterfactuals we used was simply to ask parents. Unlike many studies on kindergarten or pre-K, ours is prospective, so we asked parents at baseline what their plans were if they did not receive the scholarship. Table 10 shows the predictive power of these baseline reports on ultimate enrollment decisions for the control group. Unfortunately, these plans are much less predictive than might be hoped, possibly because parents thought we would use the information to assign scholarships, even when we truthfully explained we would not, or because these decisions are last minute.

We improve on this by following Crepon et al. (2019) in using machine learning to predict counterfactual probabilities that the child will be enrolled in kindergarten using covariates. We proceed as follows:

1. For each village, we fit a LASSO model using repeated cross validation on the other 70 villages to predict years of kindergarten and treatment status

2. Using predictors selected by LASSO we then run OLS and logit respectively to remove LASSO bias. We then use these estimations to form predictions for individuals in the selected village.

3. We repeat this 1000 times and take the median of the 1000 predictions to purge any estimation variability. As in Crepon et al. (2019), Figure 5 shows this procedure appears to successfully result in a fully converged quantity.

4. Finally, we residualize treatment with its predicted value and estimate the regression with the weights from Crepon et al. (2019) to orthogonalize the final estimate with regard to machine learning prediction error.

At the end of this procedure, we are able to examine how the treatment effect differs for those above and below the median propensity to enroll; however, if we are willing to make additional assumptions about the treatment effect distribution, the method in Hull (2018) allows us to obtain point-estimates for two different types of compliers. Specifically, we must be willing to assume: 1) that there are fundamentally two, exclusive types of compliers (those that ever go to kindergarten and those that do not without the scholarship), and 2) the subLATEs for these compliers are mean independent from our prediction of their propensity to attend kindergarten. That is, the heterogeneity of the reduced-form effects by

predicted counterfactual is due just to the predicted counterfactual shifting the shares of the types of compliers. While obviously restrictive, we believe this still adds useful information above and beyond the overall IV results.

5 Endline 1 Results

5.1 Kindergarten Enrollment

Table 4 shows that the offer of a scholarship substantially increases kindergarten attendance. Being offered the scholarship increases the likelihood of attending HLC by almost 50 percentage points, while it increases the likelihood of attending any kindergarten by around 20 percentage points. Thus compliers are roughly half those who would have not attended kindergarten in absence of the scholarship and half those who switch from other kindergartens. Interestingly, the effect is larger in the first year. This is likely because families who wish to send their children to private primary schools will enroll their children in that primary schools kindergarten in the year before first grade; during our study period, HLC specialized only in kindergarten.

5.2 Endline 1 Main Results

Column 3 of Table 5 shows being offered a scholarship increased performance on our total score index almost 0.4 standard deviations. Examining the cumulative distributions of the treatment and control group in Figure 3 shows that this is a wide-spread effect across the ability distribution. The treatment group's score distribution first order stochastically dominates the treatment group's.

These gains are also widely spread across domains. Table 7 shows that the pooled IV effect of attending HLC ranges from 0.33 to 0.97 standard deviations across domains. While it would be tempting to interpret the relative size of these domain improvements as evidence for which areas our provider is more or less skilled at teaching, it is important to remember that there are likely different elasticities of knowledge to inputs across these domains. For example, one can very directly improve mathematics performance by teaching children numbers. In contrast, improving reasoning skills requires a much more indirect approach. In addition, the effect sizes also reflect how well these skills are taught via the counterfactual activity, e.g., at the anganwadi center or at other kindergartens.

In contrast to these large improvements, we find no improvement on either the aggregate

socioemotional index nor on the subdomains of the questionnaire.The last column of Table 7 shows not only is the effect insignificant, but the point estimate is also small. One explanation for this null is that parental reports are not a good way of measuring children's socioemotional development. This concern led us to add the more direct measures to endline 2 when the children were older.

5.3 Endline 1 Heterogeneity

The prediction method outlined above in subsection 4.2 appears to yield well estimated predictions. Figure 6 shows a binscatter of the true enrollment decisions against the ones predicted by the procedure. Recall that because these predictions were formed based on the other villages we can think of this as out-of-sample performance. It achieves this fit using a variety of different predictors summarized in Table 11.

Table 12 shows how the first stage effect of the scholarship differs by propensity to enroll in Kindergarten. The first two columns show that the effect of the scholarship on inducing children to enroll in HLC is similar for those both above the median propensity to attend some form of kindergarten and below. This is in line with the first stage results shown in Table 4 that indicates a roughly equal split of compliers coming from Anganwadis and other private kindergartens. The next two columns show how the effect of the scholarship on our two exclusive fallback options differs by this estimated propensity. The estimated propensity both strongly predicts the likelihood of each fallback option among the control group and interacts with the treatment to generate variation in the shares of the different complier types.

The results of this exercise show that while all students benefit from the scholarship offer, those who switched from not going to kindergarten experience the largest gains at the first endline. The first two columns of Table 13 show the IV if enrolling at HLC estimated for those both above versus below the median of propensity to enroll in kindergarten. There is clear evidence of treatment effect heterogeneity as the effect estimated for compliers with below median propensity to attend kindergarten is significantly larger than the effect on those above the median.

The third column of Table 13 shows the estimates of the subLATEs for each complier type. The differences here are even larger. The subLATE for home compliers is three times as large as that estimated for compliers switching from other kindergartens. And this is not due to a small effect on those switching from other kindergartens. We estimate that those

students still improve performance on our tests by more than 0.4 SD.

6 Endline 2 Results

6.1 Primary Enrollment Decisions

We find no effects on subsequent enrollment decisions. The last two columns of Table 4 show children who received the scholarship are neither more likely to attend primary school at the age-appropriate time, nor are they more likely to be enrolled in a private primary school. This simplifies the interpretation of the second endline results because the only differences in education induced by the scholarship are the ECE decisions and not subsequent parental investments. Additionally, it is interesting that parents appear to neither be complimenting the educational investment induced by the scholarship nor substituting resources away.

6.2 Endline 2 Main Results

While diminished, the effect of attending HLC is still statistically and economically significant. Table 6 shows that those who were induced into attending HLC by the scholarship still score 0.4 SD better than their peers. Additionally, with the exception of the far left tail, Figure 4 shows the treatment group continues to dominate the control group's score, and Table 8 shows the gains continue to be broadly based across domains.

Interestingly we continue to find no evidence of improved socioemotional skills on any of the direct assessment measures we use or an index combining them. Table 9 shows that we find no effect across a wide variety of domains including the children's contentiousness, willingness to attempt hard problems, and their willingness to share with another child. One explanation for the lack of effect is that very few students in the control group stay home. Thus, even those in the control group who are not enrolled in Kindergarten experience very similar social environments. Anganwadi centers are similarly sized, with children of similar ages, run by women with similar characteristics.

6.3 Endline 2 Heterogeneity

Interestingly, the evidence of heterogeneity in effects is also diminished. The fourth and fifth columns of Table 13 show the pooled IV estimates for those above and below the median propensity to enroll in kindergarten. The estimated effects still different, but no longer significantly so. The subLATEs yields even less evidence of heterogeneity with almost identical point estimates. This suggests that one reason for the convergence in scores may be

that KG compliers begin to catch up on what they were missing by not being in kindergarten leaving only the relative advantage of HLC over other kindergartens.

7 Comparison of Endline 1 and Endline 2 Results

It is also useful to understand the rate at which the effects appear to be diminishing. One might be tempted to do this by simply comparing the estimated effects at the first and second endlines; however, because the test composition also changed between the two time-points and because the effects are standardized differently, this might be misleading. Thus to understand the rate of change, we create new indicies that can serve as a common benchmark by first restricting to the set of tests common across all three survey waves, and then generating new averages of z-scores standardized to the first endline control.

Using these new scores, Figure 7 shows that approximately 40% of the effect persists from endline 1 to endline 2. The natural follow-up question to this is whether the effect is diminishing because treatment students are beginning to fall short or if this is the result of control students catching up. Figure 8 gives some idea of the answer to this question by plotting the raw versions of these common scores at each of the survey time points. From this figure it appears that treatment is on a relatively constant upward trend. Conversely, there appears to have been a sharp improvement in the rate of growth for control students between the first and second endlines. This suggests that the fadeout is due to control students catching up.

8 Conclusion

As countries around the world begin to expand access to early childhood education options, there are many choices that they must make. They must decide whether it should be provided publicly or by the private market, how it should be financed, and how formal they want the instruction to be. In this paper, we provide evidence on this last question.

We find that opting for a more formal learning environment relative to play-based community centers has large and enduring effects on cognitive outcomes. Immediately after kindergarten, we find that those induced into attending a formal kindergarten have roughly doubled the rate of learning compared to their peers who did not enroll. We find that these effects are concentrated mainly on those who would not have attended kindergarten, but that there are still benefits to those who switch from other kindergarten providers. We then find that one year after the intervention, an economically important fraction of this improvement

continues to persist.

In contrast, we consistently fail to find improvements on socio-emotional learning. This is likely because the additional formality of our classrooms does not greatly change the nature of the social context. Children in both the formal kindergartens and the Anganwadi centers are organized in groups and supervised by similar authority figures. To investigate this hypothesis further it would be useful to be able to identify the effects of our scholarship on children whose counterfactual is home care. In principle, this could be done with the same machine learning approach that we use, but we unfortunately have too small of a proportion of the sample choosing home care to make this a viable strategy.

Thus, if the goal of early childhood education is to improve cognitive outcomes, there appear to be substantial benefits to investing in the infrastructure necessary to implement a more formal curriculum. Moreover, our partner demonstrates that this can be done at relatively large scale. However, if the goal is to foster socio-emotional learning, it appears unlikely that greater formality will yield additional gains beyond community pre-schools.

References

Adhvaryu, A., A. Nyshadham, T. Molina, and J. Tamayo (2018). Helping children catch up: early life shocks and the PROGRESA experiment. *NBER Working Paper*.

Almond, D. and J. Currie (2011). *Human Capital Development before Age Five*, Volume 4. Elsevier B.V.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015). Estimating the Production Function For Human Capotal: Results From a Randomized Control Trail in Colombia.

Belur, R. (2019, may). Kindergarten classes at govt schools from this year.

Berkes, J. and A. Bougen (2019). Heterogeneous Preschool Impact and Close Substitutes: Evidence from a Preschool Construction Program in Cambodia. *Mimeo*.

Blimpo, M. P. and T. Pugatch (2017). Scaling up Children's School Readiness in The Gambia: Lessons from an Experimental Study. *Mimeo*.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau (2018). Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment. *Journal of Human Resources 53*(2), 474–512.

Crepon, B., E. Duflo, E. Huillery, W. Pariente, J. Seban, and P.-A. Veillon (2019). Cream skimming and the comparison between social interventions: Evidence from entrepreneurship programs for at-risk youth in France. *Mimeo*.

Engle, P. L., L. C. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. C. De Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet 378*(9799), 1339–1353.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science 344*(6187), 998–1001.

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

Hull, P. (2018). IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons. *Mimeo*.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge: Cambridge University Press.

Jebaraj, P. (2019, jun). Draft National Education Policy proposes formal education from age of three.

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

Martinez, S., S. Naudeau, and V. Pereira (2017). Preschool and Child Development under Extreme Poverty Evidence from a Randomized Experiment in Rural Mozambique. *Mimeo* (December).

Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review 109*(4), 1426–1460.

Table 1: Child development sub-tests by domain and source

	WPPSI	DIAL	Individually Sourced
Reasoning	Matrix Reasoning	*Problem Solving*	
Memory	Picture Memory Bug Search Animal Coding		
Language	Vocabulary Receptive Vocabulary Picture Naming	*Alphabet* *Color Naming* *Action Identification*	ASER Kannada
Math		*Counting Balls*	Panamath ASER Math **Rapid Comparison** **Rapid Addition**
Creativity			Divergent Thinking
Motor Skills		*Physical Actions* *Body Parts* *Copying*	
Socioemotional			Strengths & Difficulties **Picture Big 5** **Emotion Reading** **Dictator Game** **Leadership in School**

Notes: Tests in *italics* were only administered at baseline and the first endline. Tests in **bold** were only administered at the second endline. Tests in the WPPSI column are drawn from the fourth edition of the Wechsler Preschool and Primary Scale of Intelligence. Tests in the DIAL column are drawn from the fourth edition of the Developmental Indicators for the Assessment of Learning assessment.

Table 2: Balance Test of Baseline Characteristics for Treatment and Control groups

	Treatment	Control	P-values	Normalized Differences
Child Demographics				
Age (Years)	3.975	3.995	0.503	-0.047
Girl	0.483	0.493	0.779	-0.020
Child Test Scores				
Total Score	-1.386	-1.351	0.508	-0.047
Reasoning	-0.494	-0.483	0.885	-0.010
Memory	-1.511	-1.526	0.751	0.023
Language	-1.284	-1.271	0.816	-0.017
Math	-1.292	-1.168	0.025	-0.174
Creativity	0.021	0.023	0.988	-0.001
Motor Skills	-1.127	-1.092	0.619	-0.039
Guardian Demographics				
Asset Index	0.038	-0.000	0.583	0.039
Male Education (Years)	6.739	7.212	0.098	-0.118
Female Education (Years)	7.145	7.193	0.853	-0.013
Number of children	404	404		
Joint p-value: 0.480				
Multivariate Normalized Difference: 0.255				

Notes: P-values in columns are for a test of equality between the control and treatment means. The normalized difference is the difference between the treatment mean and the control mean divided by the square root of the average variance of the sample. The joint p-value is for a test of joint equality of all listed control and treatment means. The multivariate normalized difference is computed as in Imbens and Rubin (2015).

Table 3: Test for Attrition Balance

	(1) Endline 1	(2) Endline 2	(3) Ever
Treatment	-0.00495	-0.0149	-0.0124
	(0.00852)	(0.0115)	(0.0127)
Control Mean	0.017	0.035	0.040
Observations	808	808	808

Notes: Columns show whether an individual attrits in endline 1, endline 2, or ever respectively. Attrition levels are low and uncorrelated with treatment status.

Table 4: First Stage

	(1) Ever Attended KG	(2) Ever Attended HLC	(3) Ever Attended Other KG	(4) Ever Home Care	(5) Ever Attended Anganwadi	(6) Attend Any Primary	(7) Attend Private Primary
Treatment	0.198***	0.468***	-0.286***	-0.0378*	-0.304***	0.0263	-0.0191
	(0.0303)	(0.0308)	(0.0287)	(0.0195)	(0.0323)	(0.0163)	(0.0317)
Control Mean	0.601	0.226	0.437	0.101	0.585	0.933	0.328
Observations	796	796	796	796	796	778	778

Notes: All columns control for a cubic of age, gender, and center fixed effects. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5: Main Effects on Total Score at First Endline

	(1) OLS	(2) OLS	(3) RF	(4) IV
Ever Attended HLC	0.408***	0.554***		0.836***
	(0.125)	(0.120)		(0.110)
Ever Attended Other KG	0.405***			
	(0.108)			
Ever Attended Anganwadi	-0.497***			
	(0.110)			
Treatment			0.394***	
			(0.0609)	
Control Mean	0.000	0.000	0.000	0.000
Weak IV F Stat				237.367
Observations	397	397	796	796

Notes: The outcome variable in all the four columns is the child's endline total score. The total score is calculated by normalizing various sub-tests with respect to the control, averaging the standardized values, and then re-standardizing to the control. All regressions control for baseline test score, cubic of age, gender and center fixed effects. To avoid including the effect of the scholarship, the OLS columns are restricted to the control group. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6: Main Effects on Total Score at Second Endline

	(1) OLS	(2) OLS	(3) RF	(4) IV
Ever Attended HLC	0.346**	0.309**		0.416***
	(0.141)	(0.134)		(0.134)
Ever Attended Other KG	0.503***			
	(0.136)			
Ever Attended Anganwadi	-0.176			
	(0.132)			
Treatment			0.193***	
			(0.0673)	
Control Mean	0.000	0.000	0.000	0.000
Weak IV F Stat				225.377
Observations	388	388	781	781

Notes: The outcome variable in all the four columns is the child's endline total score. The total score is calculated by normalizing various sub-tests with respect to the control, averaging the standardized values, and then re-standardizing to the control. All regressions control for baseline test score, cubic of age, gender and center fixed effects. To avoid including the effect of the scholarship, the OLS columns are restricted to the control group. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7: IV Effects by Test Domain at First Endline

	(1) Reasoning	(2) Memory	(3) Language	(4) Math	(5) Creativity	(6) Motor	(7) SEL
Ever Attended HLC	0.374***	0.551***	0.893***	0.967***	0.325**	0.430***	-0.0353
	(0.133)	(0.123)	(0.113)	(0.131)	(0.137)	(0.122)	(0.140)
Control Mean	-0.000	-0.000	-0.000	0.000	-0.000	-0.000	-0.000
Weak IV F Stat	233.471	232.201	235.612	239.162	223.357	241.239	231.588
Observations	796	796	796	796	782	796	796

Notes: The outcome variables in all the columns are standardized values of the respective test domain. The total score is calculated by normalizing various sub-tests within each test domain with respect to the control, averaging the standardized values, and then re-standardizing to the control. All regressions control for baseline test score, cubic of age, gender and center fixed effects. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8: IV Effects by Test Domain at Second Endline

	(1) Reasoning	(2) Memory	(3) Language	(4) Math	(5) Creativity	(6) SEL
Ever Attended HLC	0.242*	0.232	0.333**	0.403***	-0.00221	0.0643
	(0.143)	(0.142)	(0.136)	(0.140)	(0.142)	(0.138)
Control Mean	-0.000	-0.000	0.000	0.000	0.000	-0.000
Weak IV F Stat	222.651	220.816	223.033	226.409	221.194	221.194
Observations	781	781	781	781	781	781

Notes: The outcome variables in all the columns are standardized values of the respective test domain. The total score is calculated by normalizing various sub-tests within each test domain with respect to the control, averaging the standardized values, and then re-standardizing to the control. All regressions control for baseline test score, cubic of age, gender and center fixed effects. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 9: IV Effects by SEL Domain at Second Endline

	(1) Total Index	(2) Prosociality	(3) Conscientious	(4) Emotion Detection	(5) Motivation	(6) Leadership Position
Ever Attended HLC	0.0643	-0.00534	-0.0278	-0.0767	0.105	0.174
	(0.138)	(0.141)	(0.149)	(0.145)	(0.145)	(0.139)
Control Mean	-0.000	0.000	-0.000	0.000	-0.000	-0.000
Weak IV F Stat	221.194	221.194	221.194	221.194	220.729	219.902
Observations	781	781	781	781	773	778

Notes: The outcome variables in all the columns are standardized values of the SEL measures. The individual measure outcomes are calculated by standardizing each SEL measures with respect to the control. The total index is an average of these z-scores which is then re-standardized with respect to control. All regressions control for baseline test score, cubic of age, gender and center fixed effects. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 10: Accuracy of Control Parent Predictions

	(1) Ever Attended KG	(2) Ever Attended HLC	(3) Ever Attended Other KG	(4) Ever Home Care	(5) Ever Attended Anganwadi
Predict KG in Control	0.127**	0.0837	0.0775	-0.0112	-0.204***
	(0.0575)	(0.0511)	(0.0543)	(0.0383)	(0.0566)
Control Mean	0.601	0.226	0.437	0.101	0.585
Observations	381	381	381	381	381

Notes: All columns control for a cubic of age, gender, and center fixed effects. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 11: Predictors Chosen by LASSO

Variable Chosen	Percentage of Time Chosen
Asset Test	100.00%
Total Baseline Test Score	100.00%
Spending on Older Sibling's KG	99.92%
Female Guardian Education	99.87%
Age of Mother at First Birth	99.85%
Older Sibling Did Not Go to KG	98.34%
Average Number of Children per HH (Census)	98.05%
Travel Time to HLC	97.99%
Perceived Anganwadi Caretaker Quality	96.03%
Child is Female	95.32%
Number of Private KG in Village	84.63%
Number of Private KG in Village in Past	75.44%
Minimum You Should Pay for KG	31.87%
Baseline Picture Memory Score	5.20%
Baseline Bug Search Score	1.87%
Most You Should Pay for KG	1.58%
Parent Time Working Away from Home	1.53%
Unemployment Rate (Census)	0.89%
Perceived Other KG Facility Quality	0.36%
Percentage Scheduled Caste Women (Census)	0.23%
Percentage Scheduled Caste Men (Census)	0.20%
Perceived Anganwadi Facility Quality	0.07%
Percentage Scheduled Caste (Census)	0.06%
Baseline Vocabulary Score	0.03%

Notes: Items appear in this table if any form of this variable (e.g. square, levels, indicators for levels of ordinal variables) are selected by LASSO. Percentages of time chosen by LASSO are computed weighting each of the 1000 iterations and each of the 71 villages equally. Variables never chosen are omitted for space.

Table 12: First Stage By Propensity to Enroll in KG

	(1) Attend HLC (Below Median)	(2) Attend HLC (Above Median)	(3) Always Home Care or Anganwadi	(4) Ever Attended Other KG
Treatment	0.509***	0.475***	-0.417***	0.0194
	(0.0437)	(0.0459)	(0.0710)	(0.0640)
Treatment × Predicted Years KG			0.229***	-0.317***
			(0.0609)	(0.0623)
Predicted Years KG			-0.223***	0.144***
			(0.0418)	(0.0376)
Control Mean	0.161	0.288	0.399	0.437
Observations	396	400	796	796

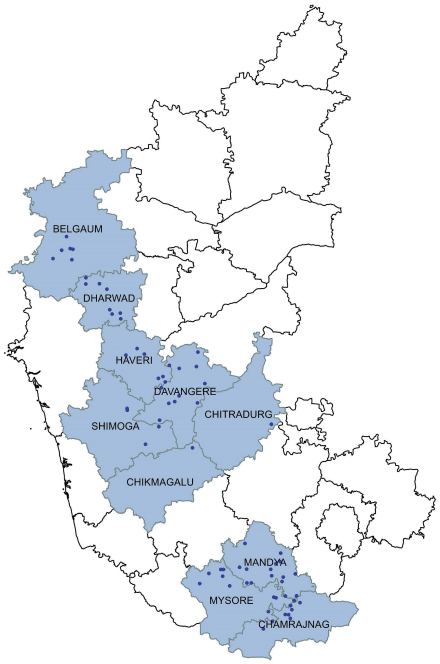
Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 13: Effect on Total Score by Propensity to Enroll

	Endline 1			Endline 2		
	(1) Below Median	(2) Above Median	(3) subLATE	(4) Below Median	(5) Above Median	(6) subLATE
Ever Attended HLC	1.030***	0.622***		0.539***	0.306	
	(0.149)	(0.147)		(0.175)	(0.187)	
Home Complier			1.351***			0.324
			(0.293)			(0.402)
KG Complier			0.440**			0.492
			(0.185)			(0.302)
Control Mean	-0.396	0.375	0.000	-0.292	0.282	-0.000
Equal Effects P-Value		0.033	0.035		0.281	0.796
KP F-Stat	135.631	107.287	12.286	123.833	100.725	9.135
Observations	396	400	796	394	387	781

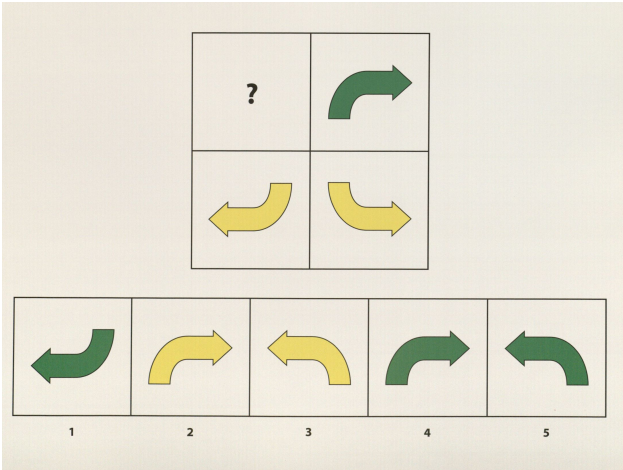
Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure 1: Study Villages



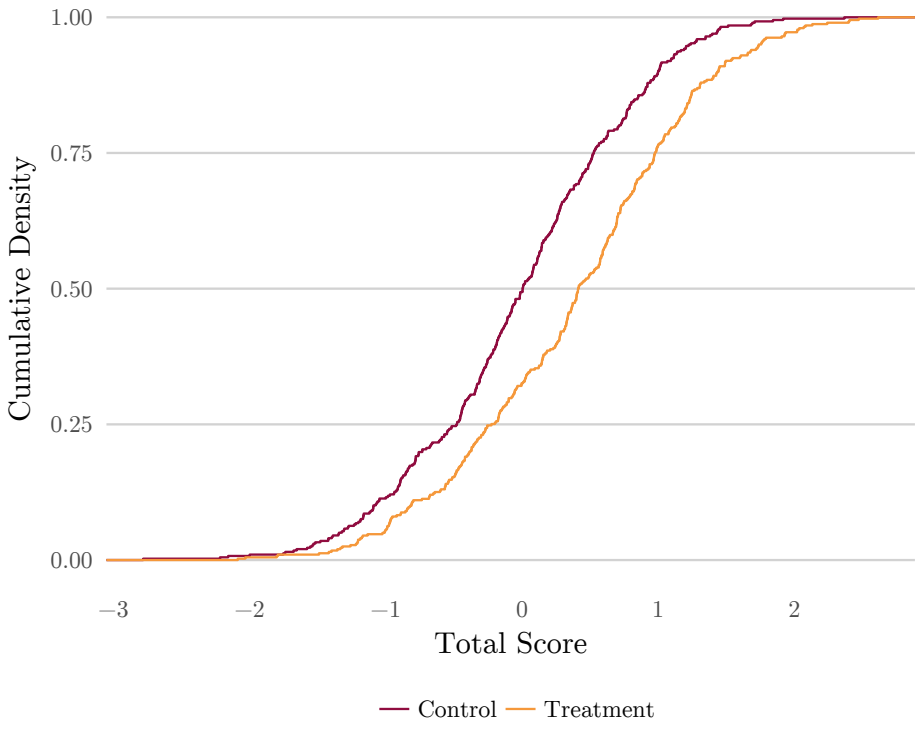
Notes: This map shows the locations of the study villages in dark blue and the ten districts of Karnataka in which the villages are located in light blue.

Figure 2: Example of reasoning assessment



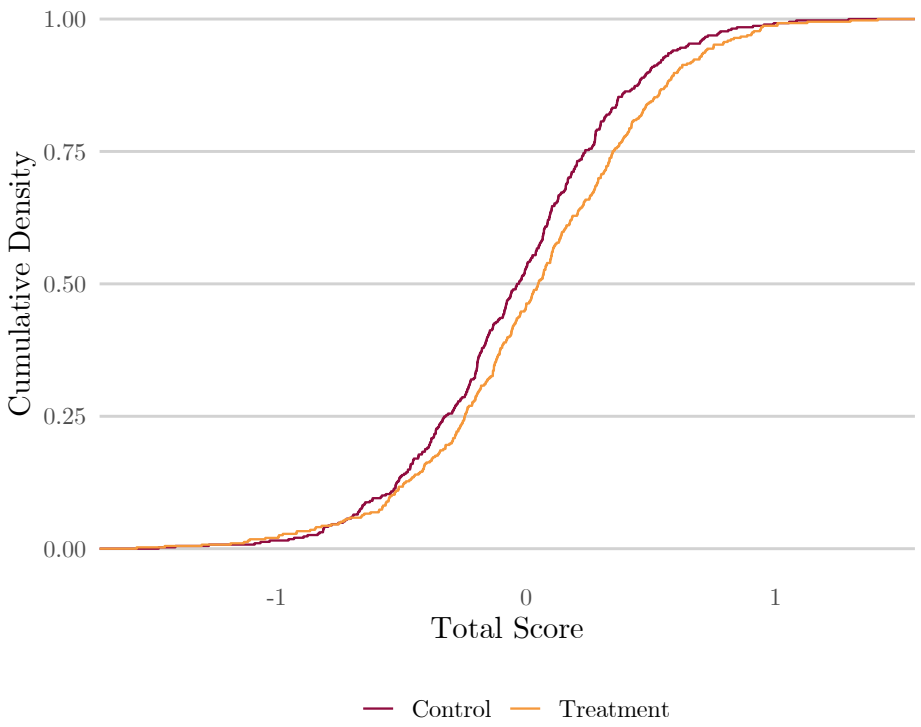
Notes: This measure assess children's ability to use reasoning to complete patterns. They are shown a series of boxes with a space missing and have to choose from among the available options which correctly completes the pattern. The test begins with very simple patterns and proceeds to become more difficult until children incorrectly answer three questions in a row. At no time are children given feedback on their performance.

Figure 3: Effects of Scholarships on Total Score Distributions at Endline 1



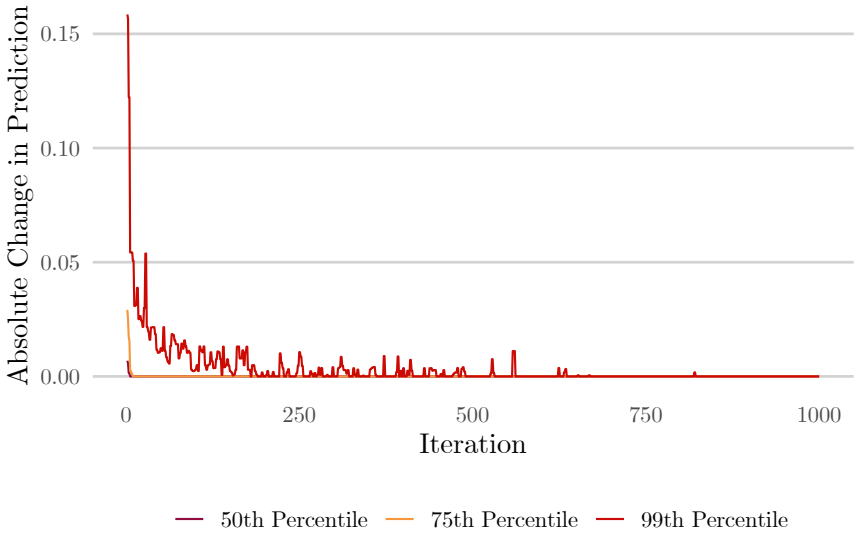
Notes: This figure shows the CDFs of the total score variable for both the treatment and control groups at the first endline. The scores of those that received a scholarship offer first order stochastically dominate those of the control.

Figure 4: Effects of Scholarships on Total Score Distributions at Endline 2



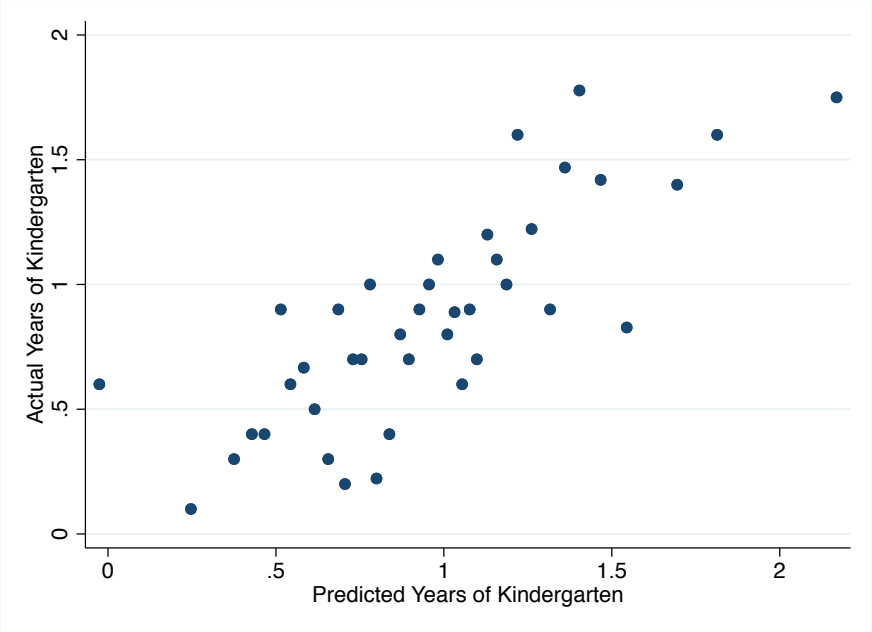
Notes: This figure shows the CDFs of the total score variable for both the treatment and control groups at the second endline. The scores of those that received a scholarship offer first order stochastically dominate those of the control except at the left tail.

Figure 5: Convergence of ML Predictions Over Iterations



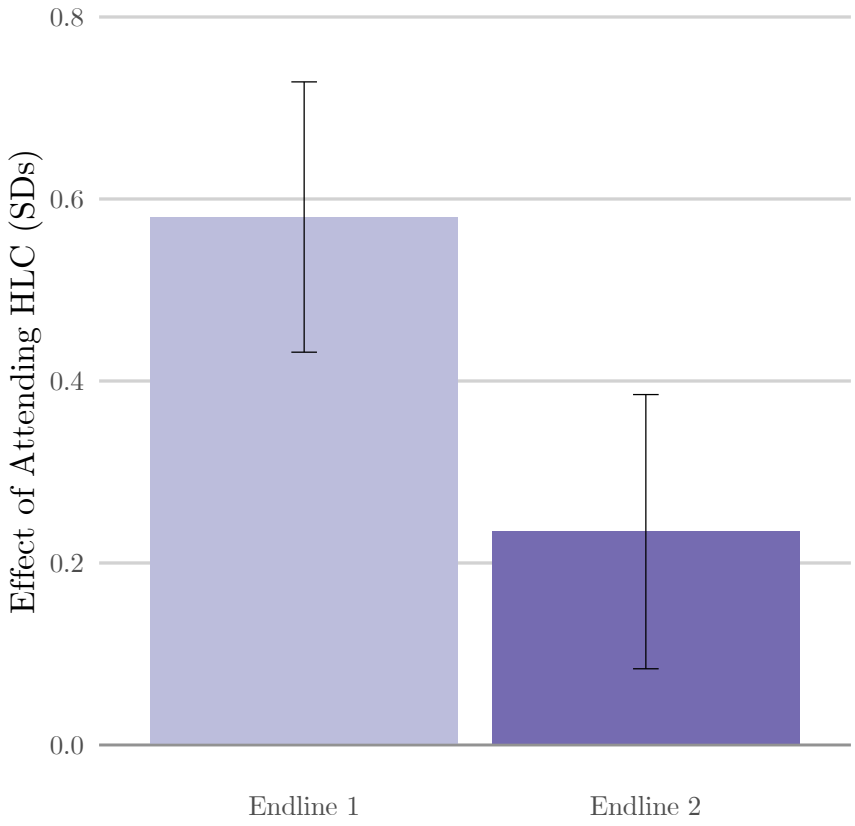
Notes: This figure shows how the prediction of years of kindergarten changes over each iteration of the procedure. To better understand the convergence, each line shows a different percentile of change at each iteration. By the 1000th iteration, even the 99th percentile change is zero.

Figure 6: Binscatter of Predicted Years of Kindergarten and Actual Years of Kindergarten



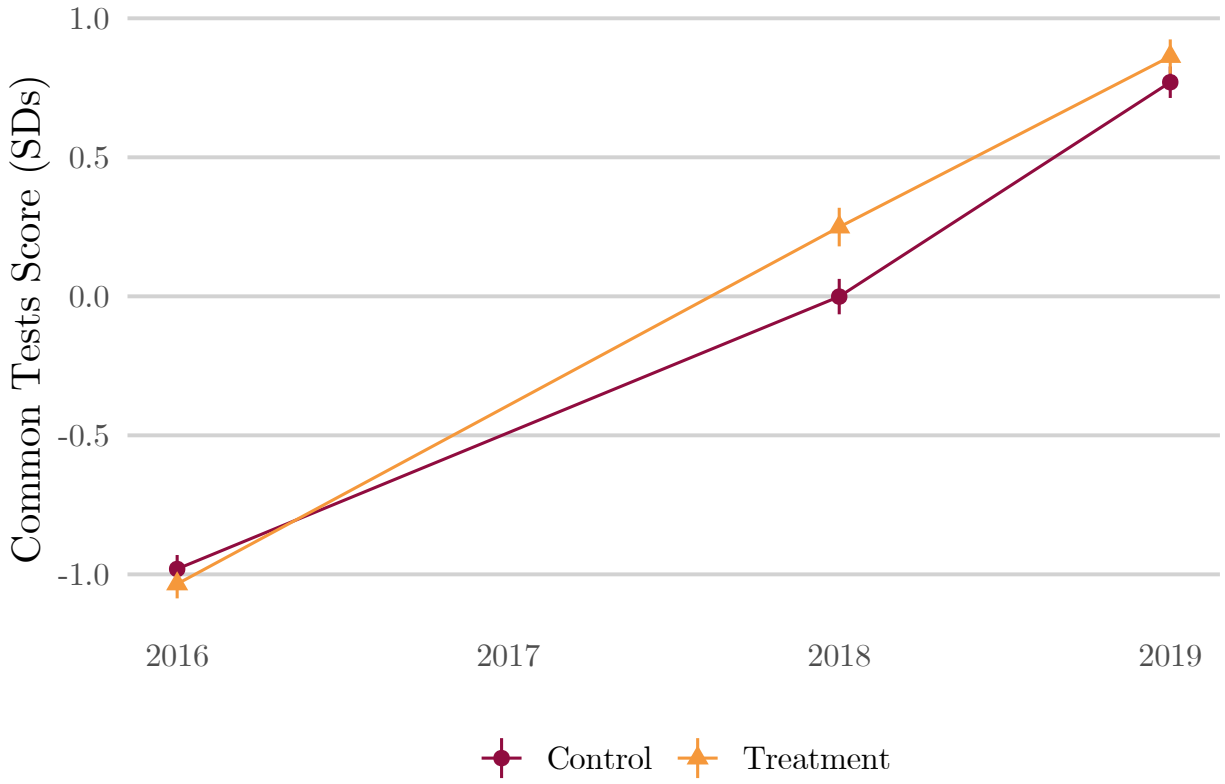
Notes: This figure shows a binscatter of the actual years of kindergarten against the years predicted by the machine learning procedure. Because each prediction is formed with models trained on the other villages, these predictions can be considered out-of-sample.

Figure 7: Comparison of Pooled IV on Common Tests by Endline



Notes: This figure shows the pooled IV treatment effect estimated on on a total index of the tests common to all time points, after being standardized to the first endline control to facilitate comparison of effect sizes over time. The bars show that approximately 40% of the effect persists from the first to the second endline.

Figure 8: Raw Scores on Common Tests Over Time



Notes: This figure shows the raw score on a total index of tests common to all time points, after being standardized to the first endline control to facilitate comparison of effect sizes over time. The trends appear to show treatment on a stable upward trajectory while control improves the rate of learning after 2018 when they enter primary school.