A Randomized Evaluation of an On-Site Training for Kindergarten Teachers in Rural Thailand*

Weerachart T. Kilenthong, Sartja Duangchaiyoosook,
Wasinee Jantorn and Varunee Khruapradit
Research Institute for Policy Evaluation and Design (RIPED)
University of the Thai Chamber of Commerce

Abstract

This study evaluates the effectiveness of intensive and hands-on on-site training for preschool teachers using a randomized controlled trial in rural Thailand. The main finding is that the intervention led to an increase in the effectiveness of the classroom in terms of children's cognitive skills by almost 50 percent relative to the control group. The on-site training intervention is cost-effective, costing 32.7 USD per student. Further investigation reveals that its specificity regarding the teaching approach or curriculum and detailed weekly teaching plans could be critical to its success.

Keyword: teacher training, teacher professional development, early childhood, school readiness, on-site training, randomized controlled trial

JEL Code: I21; I25; J24.

1 Introduction

A high-quality early childhood education (ECE) program can be an effective tool to foster child development and promote human capital accumulation (e.g., Elango et al., 2015; Heckman and Masterov, 2007). There are several early childhood curricula or approaches that have shown promise in small-scale experiments in developed countries (e.g., Campbell et al., 2002; Heckman et al., 2010; Schweinhart and Weikart, 1997).

*The authors gratefully acknowledge helpful comments from Fernando Andrade-Adaniya, Suparee Boonmanunt, Flavio Cunha, James Heckman, Kraiyos Patrawart, seminar participants at the HCED of the University of Chicago, and participants at AASLE Conference 2024. We also want to thank the early childhood experts who tirelessly implemented the intervention, especially Khanista Boonsanong, Rampai Chaipalee, Kittisak Ketunuti, Suthara Maliya, Rattanaluck Maneerattanachotiwong, Manee Panjanghan, Anchalee Sittikultorn and Warabud Suppalarkbunlue, and the RIPED data collection team who made all the data available. We are grateful for financial support from the Equitable Education Fund (EEF). The findings and conclusions contained in the report are those of the author(s) and do not necessarily represent the views of the funders. Corresponding author: Weerachart T. Kilenthong. Email: tee@riped.utcc.ac.th (Weerachart T. Kilenthong); kei@riped.utcc.ac.th (Sartja Duangchaiyoosook); wasinee_jun@riped.utcc.ac.th (Wasinee Jantorn); varunee@riped.utcc.ac.th (Varunee Khruapradit).

However, these proven programs have not yet been adopted in rural areas of developing countries. The question is: why not?

This study tests whether a well-proven program, innovated in a developed country like HighScope, can be adopted effectively in the rural setting of a developing country, Thailand. This is not the first attempt in Thailand. Chujan and Kilenthong (2021) evaluated an intervention to promote a HighScope-based curriculum by randomly assigning additional teachers. They found that the intervention positively and significantly affected child development in gross motor and personal and social skills. However, hiring an additional teacher is costly since most childcare centers in Thailand already have a sufficient number of teachers. That insight led to the development of a hands-on, on-site training to scale up. More specifically, this paper evaluates whether on-site teacher training can be a teacher's professional development to improve child development effectively.

The closest paper is Andrew et al. (2022), which evaluated the impact of teacher training in Colombia. They found that online teacher training, regular support, and a complete set of tasks or content consistent with the curriculum can be effective. Their experiment has two treatments: additional teacher and teacher training. It is unclear whether teacher training alone could be as effective as a combination. On the other hand, our study has only one treatment, on-site training. Therefore, our result can be directly attributed to teacher training. Similar to Andrew et al. (2022), we observed teaching quality based on school visit records by an expert team. Therefore, we can show that the intervention has significantly improved teaching quality, especially in the Plan-Do-review process and overall classroom quality.

This study is also closely related to Loyalka et al. (2019), who evaluated a large-scale junior high school teacher training program in China. The paper found that the program could not improve teacher and student outcomes after one year of the intervention, and one potential explanation is that the training content was overly theoretical. On the contrary, our training was hands-on and highly specific regarding the teaching approach or curriculum. It also provided the trainees with detailed weekly teaching plans covering all daily activities (consistent with the conclusion in Andrew et al., 2022; Banerjee et al., 2007). These differences potentially explain why we found significant impacts on children's outcomes and why Loyalka et al. (2019) could not. Another potential reason could be the difference in the target groups, i.e., kindergartners versus junior high schoolers.

The remainder of this paper is organized as follows. Section 2 describes the experimental design of the intervention. In section 3, we explain how the data were collected. Section 4 presents the benchmark results and proposes potential mechanisms, while section 5 extends the analysis to account for the concavity of skill production function and skill depreciation. The heterogeneous effects and robustness checks are presented in section 6 and 7, respectively. Section 8 concludes the paper and provides further discussion.

2 Experimental Design

2.1 Preschool and Kindergarten in Thailand Context

Most young children in rural Thailand usually start their preschools between two and three years old in a childcare center in their community. They would stay there for two years. They would then move to a nearby school to enter kindergarten for two years before entering Primary one, Thailand's first formal schooling year. This study was exclusively conducted in kindergarten classes of rural schools in Roi-Et province.

2.2 On-Site Teacher Training as the Treatment

This experiment has only one treatment: on-site teacher training in real classrooms. This training has been designed and operated by the Reducing Inequality through Early Childhood Education (RIECE) program, which aims to improve early childhood education in rural Thailand. The main objective of the training is to enable the trainees to implement a HighScope-based curriculum, called the RIECE curriculum, in their classrooms. The trainees/teachers have to learn and practice with real students in early childhood classrooms of one of the training centers of the RIECE program, which is Muang Roi-Et school in Roi-Et province, in this experiment.

It is worth emphasizing that this training differs from regular on-the-job training, usually performed in the trainee's classroom, in that it is staged at a training center when the class is in session. The main reason for having the training at a training center instead of the trainee's classroom is that teachers from rural Thailand tend to imitate what they have seen and experienced during the training rather than capture key principles and learn how to apply them in their classrooms. Therefore, a training center must have a sufficiently high quality.

This is intensive and hands-on training, two weeks (10 school days). The main task for the first three days is to observe all classroom activities and student behaviors. The rest is for the trainees to practice implementing all key activities and producing learning materials. The training strongly emphasized how to implement the plan-do-review process (PDR), which is the core concept of the HighScope approach. In addition, there will be a review session every afternoon, where the trainees, the homeroom teachers, and the trainer will reflect and discuss how well each trainee performed each activity during the day. This session also introduces relevant theoretical concepts and discusses their connection with the daily experiences of the trainees. Note that the intensity of the training limits the training capacity. As a result, 37 treated teachers were trained in three separate rounds at different points in time (see the timeline in figure 1 for details).

2.3 In-Class Training and School Visits for both Control and Treatment Schools

All participating teachers at the baseline, from both control and treatment, were invited to a two-day in-class training, which focuses on theoretical concepts of the curriculum, including active learning, adult-child

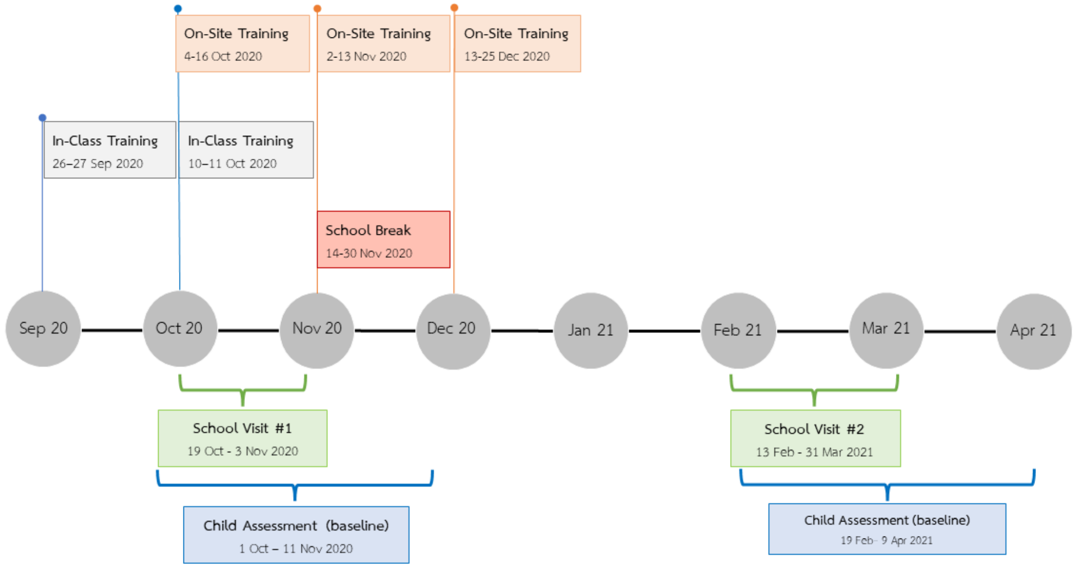


Figure 1: Timeline of the experiment.

interaction, learning environment, daily routine, assessment, and key developmental indicators (KDIs). The training was mainly theoretical, with short practice sessions. Importantly, this training involves no real students and is staged at a convention center, not an early childhood classroom. One hundred thirty-two teachers participated in one of the two training sessions, and 115 completed the training.

In addition, an early childhood education expert visited each participating classroom in both control and treatment groups twice, at the baseline and the endline. The visit had two main purposes: advising teachers on implementing the curriculum and collecting learning quality information. This was an intensive task, which took a whole day for each school. As a result, we had to employ eight experts/evaluators. We account for potential measurement errors due to possibly heterogeneous scales of different visitors using evaluator-fixed effects in every estimation involving learning quality.

2.4 Randomization

Only teachers who completed the on-site training were eligible to be randomized. The randomization was done twice after each in-class training session was completed. We had to drop extremely small schools whose early childhood students were less than five. There were 67 eligible schools. However, two did not allow school visits, so they were excluded. As a result, 65 schools participated in the experiment. The randomization unit is a school. The treatment and the control groups comprised 29 and 36 schools, respectively. The program randomly invited 29 schools to send their early childhood teachers to the on-site training, and 28 complied.

In the end, 38 teachers from 28 schools attended the training.

3 Data and Measurements

This study directly assessed students twice, at the baseline and the endline, using the school readiness survey instruments[1]. The school readiness instruments assess the students' language/literacy, mathematics/numeracy, fine motor, and working memory. See the online appendix for the assessment tools. In addition, the survey includes questionnaires for both teachers and parents to evaluate the social-emotional skills and personal characteristics of the students, while the parent questionnaire collects parents' characteristics and household socio-economic status as well. Unfortunately, with limited resources, both questionnaires were asked only at the endline, and all were self-completed.

The baseline and endline surveys assessed 1,077 and 998 students, respectively (attrition rate of 7.3%). The attrition rates for the control and treatment groups were 6.5% and 8.2%, respectively. The attrition is equally distributed across both groups. We, therefore, should not be concerned about the attrition issue.

There were two potential contaminations during the experiment. First, teachers from two out of 29 schools in the treatment groups participated in the on-site training a few years earlier. Second, 88 students in the treatment groups were assessed after their teachers received the treatment (completed the on-site training). We dropped students from both cases from the main sample. As a result, our main sample includes 866 students and 75 teachers from 70 classrooms in 57 schools (21 treatment and 36 control). We also present the estimation results using the whole sample of 998 students (29 treatment and 36 control) and 88 teachers from 81 classrooms in 65 schools (balanced panel sample) as a robustness check.

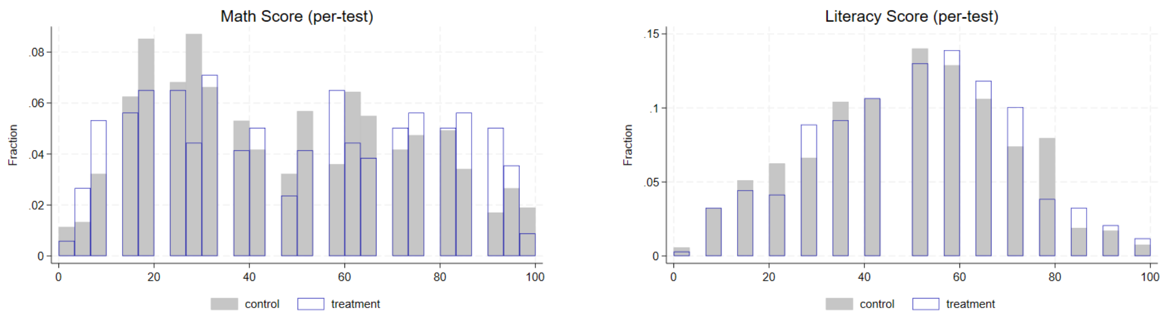
3.1 Child Cognitive Skills

Child cognitive skill is derived from mathematics/numeracy and literacy/language items.[2] There are 35 items in total: 21 for mathematics and 14 for literacy. Our main measure for child cognitive skills, cognitive test score, is the percent of items answered correctly by the child, i.e., $\theta_i^c = \frac{C_i}{35} \times 100$ where C_i is the number of cognitive items child i answered correctly. Similarly, measures for math and literacy skills are $\theta_i^m = \frac{M_i}{21} \times 100$ and $\theta_i^l = \frac{L_i}{14} \times 100$, where M_i and L_i are the numbers of math and literacy items child i answered correctly, respectively. We use these raw scores as our main measures because they are simple and appropriate for estimating daily learning gain, our main outcome in this paper. We also performed the estimations using item response theory scores (IRT scores) as the outcomes for robustness checks. Figure 2

[1]This assessment is adapted from the Measure of Development and Early Learning (MODEL) under the Measuring Early Learning and Quality and Outcomes (MELQO), which is a collaboration of UNESCO, World Bank, Brookings Institution and UNICEF (UNICEF, 2012). See Kilenthong et al. (2023) for more details.

[2]We do not use working memory test scores because less than 25 percent of the sample could perform the test in baseline and endline surveys. A student could perform the digit span memory task only if he/she passed the number identification test with full scores (knowing all one-digit numbers).

presents the distributions of math and literacy test scores at the baseline. Note that the test scores of the treatment and control groups are relatively similar. See table A.1 for the formal balanced tests.



(a) Math test scores at the baseline.

(b) Literacy test scores at the baseline.

Figure 2: Math and literacy test scores at the baseline for treatment and control groups.

3.2 School Days

We measure school days for each child using the number of potential school days between the baseline and endline tests. This excludes weekends, official holidays, and the days her teacher(s) had been absent when attending the on-site training. Unfortunately, we cannot exclude the days a specific child was individually absent from class because it was unobserved.

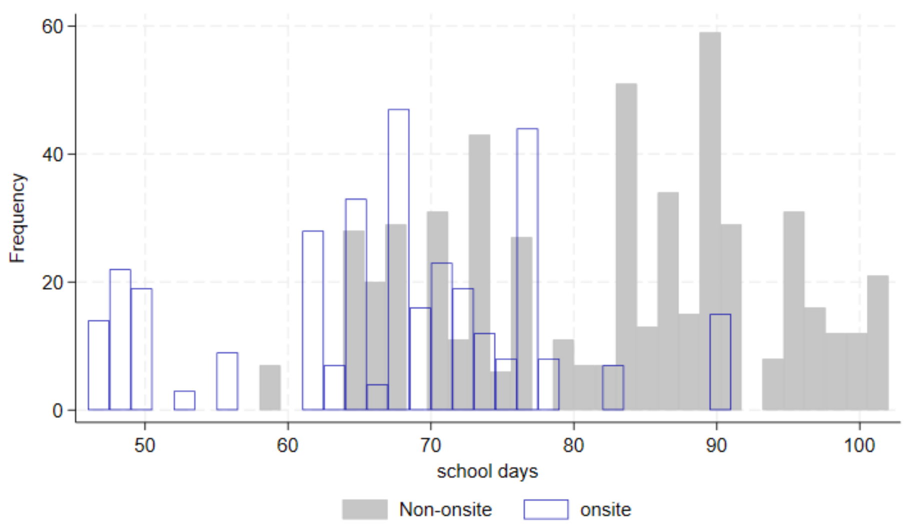


Figure 3: Histograms of school days between the baseline and endline tests.

Direct assessment for young students is an intensive and time-consuming process. An average assessment time was about 18 minutes per child and was naturally carried out for one student at a time. As a result, it took at least one day to complete the assessment for each school. With a limited number of research staff (6 persons), the baseline and endline assessments took 29 and 35 school days, respectively, to complete. Ideally, we could have scheduled the endline test so that each student had a similar number of school days between the two tests. Unfortunately, that was not the case. The difference between the maximum and the minimum number of school days is more than 50. See figure 3. The figure also reveals that the distributions of school days for the control and treatment groups are distinct. This difference results from both the testing schedule issue and the fact that treated teachers had to leave her classroom for almost two weeks.

To account for these differences in school days, we use learning gain per school day, called daily learning gain, as our main outcomes. See the next section for more specific details.

3.3 Summary Statistics and Balanced Tests

Panel A and B of table A.1 present key variables from baseline and endline surveys, respectively. For comparison, the results are categorized into treatment and control groups. For brevity, we comment on some variables only. The sample was gender-balanced; that is, 49.3 percent of the sample was female. The average age of sampled students was about 5.54 years old, while the average student-teacher ratio was roughly 14.09. Most children lived in relatively large households with an average household size of 5.78 and relatively low-educated caregivers (only 26 percent with high school or above). Their home environments were relatively poor, with less than three children's books per family, and average parents read to their children less than two days per week. Overall, the majority of children were relatively disadvantaged. On the other hand, their teachers were well-educated, with more than 90 percent holding a bachelor's degree and 74 percent majoring in early childhood education (ECE).

The last two columns of table A.1 present the results of balanced tests, where we regressed each variable on the treatment dummy (and a constant), and standard errors were clustered at the school level. All but two variables are not significantly correlated with the treatment, except the fraction of teachers holding a bachelor's degree and the fraction of teachers attending the in-class training. Importantly, all key variables from the baseline were not significant, especially the math and literacy scores. The overall results indicate that the randomization is valid. The same conclusion can be drawn from the whole sample as in table A.2 in the appendix.

4 Impact of On-Site Training

4.1 Benchmark Model

Outcome variables for our benchmark models are in the form of daily learning gain, measured by the value-added test scores divided by the number of school days between the baseline and endline tests. The daily learning gain captures the effectiveness of the classroom for each skill.

More formally, let θ^s_{i0} and θ^s_{i1} denote test scores of student i for skill s at the baseline and the endline, respectively. The daily learning gain of student i for skill s is defined by $\frac{\theta^s_{i1}-\theta^s_{i0}}{\tau_i}$, where τ_i is the number of school days of student i.

The benchmark model estimates the following linear regression model:

$$\frac{\theta^s_{i1}-\theta^s_{i0}}{\tau_i} = \alpha^s + \beta^s T_i + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon^s_i, \tag{1}$$

where T_i is a dummy variable indicating whether student i attended a treatment school during the experiment, $\boldsymbol{X}_i$ is a vector of control variables[3], and ε^s_i is an error term. The key parameter of interest, β^s, estimates the intent-to-treat effect (ITT) of the on-site teacher training on student's skill s.

There were a couple of non-compliance issues in this experiment. Some teachers who attended the on-site training were later assigned to teach in other classrooms or schools, while some others in the treated classrooms did not attend the on-site training. To deal with the non-compliance problem, we estimate the following treatment-on-the-treated effect (TOT) using the treatment dummy T_i as the instrument.

$$\frac{\theta^s_{i1}-\theta^s_{i0}}{\tau_i} = \alpha^s + \beta^s A_i + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon^s_i, \tag{2}$$

where A_i is the fraction of teachers in student i's classroom who completed the on-site training and is instrumented by the treatment dummy T_i.

4.2 Benchmark Results

Panel A of table 1 presents the intent-to-treat effects (ITT) of the on-site teacher training on cognitive (the sum of math and literacy), math, and literacy skills. The results imply that the intervention significantly impacts children's skills. In particular, randomly assigning teachers to the on-site training can significantly improve the daily learning gain of young students in all three skills.

For the basic model with the minimum controls (maximum sample), on-site teacher training can raise students' daily learning gains for cognitive, math, and literacy skills by 0.0664, 0.0578, and 0.0795, respectively, which are equivalent to 49, 39, and 69 percents of the average corresponding daily learning gains of the control group (roughly 0.1347, 0.1482, and 0.1146, respectively). See figure 4 for an illustration. Given that the average school days for the treatment group were 67 days, the results suggest that the total impacts of the interventions on cognitive, math, and literacy skills are 4.45, 3.87, and 5.33 (out of 100), respectively.

We can transform the impacts in terms of baseline standard deviations of the control groups as follows. For the basic model, the impacts on cognitive, math, and literacy skills are approximately 0.00302, 0.00222,

[3] This paper primarily uses two sets of control variables. The first set, $\boldsymbol{X}^1$, includes age, age squared, child weight, child height, child gender, student-teacher ratio, the fraction of teachers who attended the in-class training, and grade level (K2 class, K3 class or mixed class). The second set, $\boldsymbol{X}^2$, includes all variables in $\boldsymbol{X}^1$ and more teacher and household characteristics, including a special-need child dummy, main language at school (standard Thai dummy), average teacher age, a fraction of teachers with a bachelor degree or above, a fraction of teachers with an ECE degree, living in the urban area, wealth, household size, caregiver's education, main language at home (standard Thai dummy), and the number of days adults read to the child. Note that adding more variables in $\boldsymbol{X}^2$ leads to a significant drop in the sample (roughly from 866 to 551).

Table 1: Estimation results of the intent-to-treat effects (ITT) and the treatment-on-the-treated effects (TOT) for the benchmark model.

	COG	COG	MATH	MATH	LIT	LIT
Panel A: ITT with Raw Scores						
Treat	0.0664***	0.0804***	0.0578**	0.0775***	0.0795***	0.0848***
	(0.0208)	(0.0202)	(0.0221)	(0.0246)	(0.0256)	(0.0268)
N	866	552	866	552	866	552
Panel B: ITT with Standardized Scores						
Treat	0.00238**	0.00347***	0.00163*	0.00285***	0.00313**	0.00361***
	(0.00103)	(0.00101)	(0.000933)	(0.00106)	(0.00124)	(0.00127)
N	865	551	864	551	864	551
Panel C: TOT with Raw Scores						
Compliance	0.0748***	0.0910***	0.0650**	0.0877***	0.0895***	0.0960***
	(0.0260)	(0.0254)	(0.0268)	(0.0301)	(0.0315)	(0.0313)
N	866	552	866	552	866	552
F-Stat	252.9	253.5	252.9	253.5	252.9	253.5
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$

Note: Clustered-standard errors at the school level are in parentheses: * $p<0.05$, ** $p<0.01$,*** $p<0.001$. $\boldsymbol{X}^1$ includes age, age squared, child gender, child weight, child height, student-teacher ratio, the fraction of teachers who attended the in-class training, and grade level. $\boldsymbol{X}^2$ includes all variables in $\boldsymbol{X}^1$ and more teacher and household characteristics, including a special-need child dummy, main language at school, average teacher age, a fraction of teacher with a bachelor degree or above, a fraction of teacher with an ECE degree, living in the urban area, wealth, household size, main caregiver's education, main language at home, and days adults read to the child.

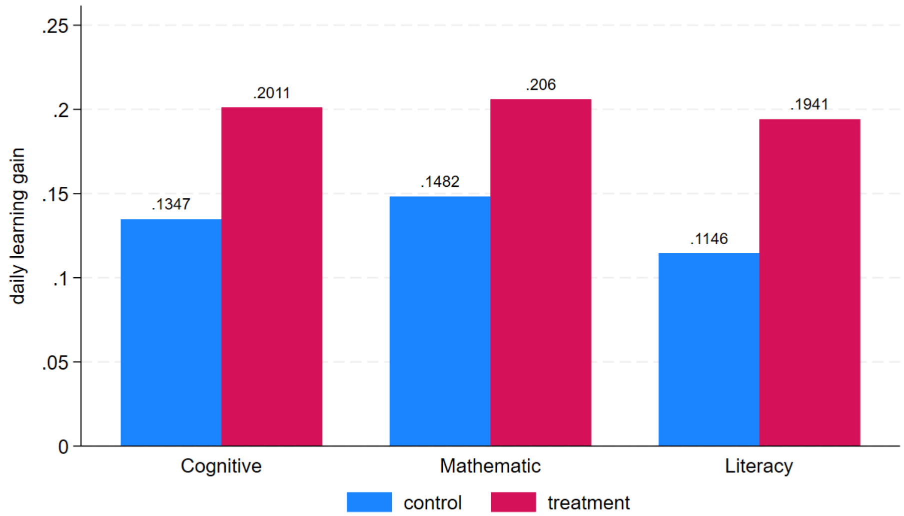


Figure 4: The intent-to-treat effects (ITT) of the on-site teacher training on children skills.

and 0.00375, the control group standard deviations (SD) of the corresponding skills at the baseline, respectively. Alternatively, we can re-estimate the models where the daily learning gains were calculated from the corresponding age-standardized scores.[4] The results are reported in panel B of table 1. The impacts on cognitive, math, and literacy skills are about 0.00238, 0.00163, and 0.00313 standard deviations (SD), respectively. These numbers are comparable but slightly smaller than the transformed version. Again, given that the average school days for the treatment group were 67 days, the results suggest that the total impacts of the interventions on cognitive, math, and literacy skills are 0.16, 0.11, and 0.21 standard deviations (SD), respectively. With further interpolation, the impacts of the intervention on cognitive, math, and literacy skills, if implemented one academic year (200 school days), should be approximately 0.48, 0.33, and 0.63 standard deviations (SD), respectively.

Panel C of table 1 presents the treatment-on-the-treated effects (TOT) of the on-site teacher training on cognitive, math, and literacy skills. The estimation results are remarkably close to the ITT effects in panel A. For example, the TOT effect on the cognitive skills for the basic model is about 0.0748 compared to 0.0664 for the ITT. This inappreciable increase is expected since the compliance rate is very high at 94.4% (based on classroom data) and 91.6% (based on student data).

We also perform a similar estimation with child weight and height to check if the above results could be spurious since the training has nothing to do with nutrition, potentially affecting weight and height. The results in table 2 show that the intervention has no significant impact on daily gains in weight and height. This indicates that it is likely that the above results are not spurious.

[4] Following Attanasio et al. (2020), we derived age-standardized scores for all skills using kernel-weighted local polynomial smoothing up to the third-degree polynomial. Few observations were dropped because their predicted variances were negative.

Table 2: Estimation results for the impacts of PDR and overall quality indices on daily learning gains using the instrumental variable approach.

	Weight	Weight	Height	Height
Treat	0.0340	-0.00967	-0.00499	-0.102
	(0.278)	(0.328)	(0.435)	(0.426)
N	866	552	866	552
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

4.3 Potential Mechanisms

This paper considers two main channels through which on-site training can potentially impact students' learning, namely learning quality in the classroom and home environment, using the information collected during the two school visits and the parent questionnaire.

An early childhood expert observed and collected learning-quality-related data during each visit based on a classroom observation record. The research team designed this classroom observation record to capture the key concepts of the curriculum, including the learning environment in the classroom, daily routine (plan-do-review, large-group time, small-group time), teacher-child interaction, teacher supports, teacher preparation/planning, school environment, and school management. We employed exploratory factor analysis techniques (EFA) to generate learning-quality indices. The analysis suggests that there are two factors for classroom data, three for teacher data, and two for school data.[5] See the online appendix for more details. We then formulated the dedicated measurement system or factor model in which each item only proxies one latent factor and estimated the factor score for each latent factor separately using the Bartlett method (Bartlett, 1937). Based on the content of items dedicated to each factor, we named those seven factors/indices as follows. The two factors from classroom data are the plan-do-review quality index (PDR) and the overall-classroom quality index (Overall). The three factors from teacher data are the teacher-child interaction quality index (Interaction), the teacher-support quality index (Supporting), and the teacher-preparation quality index (Preparation). The last two factors from school data are the school-environment quality index (Environment) and the school-management quality index (Management).

For the basic home environment, there are three variables, namely the number of children's books at home (No. Books), the number of days parents read to the child (Days Read), and the dummy indicating whether the child has been playing games on tablet/cellphone/computer during the last week (Playing Tablet). These variables are supposed to capture parental investments and household responses to the intervention.

[5]We determine the number of latent factors for each group using the eigenvalue criteria (eigenvalue is larger than one) and factor loadings after performing factor rotation with quartimin rotation. In particular, we will disregard an item whose factor loading after the quartimin rotation is less than 0.3, following Attanasio et al. (2020).

We estimated the intent-to-treat effects (ITT) of the on-site teacher training on learning quality using the following linear regression.

$$Q_j = \alpha + \beta T_j + \boldsymbol{\gamma} \boldsymbol{Z}_j + \varepsilon_j, \tag{3}$$

where Q_j is a quality index of unit j (classroom, teacher, or school)[6], T_j is a dummy variable indicating whether unit j is in the treatment group, and $\boldsymbol{Z}_j$ is a vector of control variables[7] for unit j.

Table 3 presents the estimation results of the model (3) for both before and after the intervention. The results in panel A (before the intervention) indicate that the learning quality of treated and controlled classrooms/schools was similar (at least in the statistical sense). If anything, treated students had poorer learning quality in all dimensions before the intervention.

Table 3: The estimation results of the model (3) for both before and after the intervention.

	PDR	Overall	Interac.	Supp.	Prep.	Envi.	Manage.
Panel A: For the Baseline Indices with $\boldsymbol{Z}^0$ as Controls							
Treat	-0.135	-0.114	-0.0811	-0.252	-0.0984	-0.0752	-0.0465
	(0.205)	(0.302)	(0.246)	(0.237)	(0.230)	(0.279)	(0.287)
N	69	69	70	70	70	70	70
Panel B: For the Endline Indices with $\boldsymbol{Z}^1$ as Controls							
Treat	0.877***	0.522**	0.283	0.271	0.278	0.278	0.240
	(0.287)	(0.211)	(0.209)	(0.210)	(0.333)	(0.226)	(0.252)
N	69	69	68	68	68	69	69

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001.

On the other hand, the results in panel B (after the intervention) show that the intervention significantly improved the plan-do-review (PDR) and the overall-classroom (Overall) quality index. In particular, after the intervention, treated classrooms had higher quality regarding the PDR and the overall indices. Unfortunately, we found no significant impact on the teacher-child interaction (Interac.) or the teacher-support (Supp.) quality index, even though they were also extensively emphasized during on-site training. One possible explanation is that both concepts are abstract and, therefore, are not easy to implement in practice. On the other hand, the PDR and the overall quality indices are based on specific activities with detailed manuals; therefore, they are relatively easier to implement. The results of the school-level quality indices suggest that the intervention did not affect the school environment (Envi.) and the school management (Manage.) quality index. This is as expected since the intervention focused on the classroom level only.

[6] A teacher-related quality index is an average score of all teachers in the classroom. That is, the unit for all three teacher-related indices is a classroom.

[7] The controls for the baseline scores, $\boldsymbol{Z}^0$, include student-teacher ratio, teacher age, teacher education, a fraction of teacher with an ECE degree, grade level, evaluator-fixed effects. The controls for the endline scores, $\boldsymbol{Z}^1$, include a fraction of teachers who attended in-class training, all quality indices at the baseline, and all variables in $\boldsymbol{Z}^0$.

This study found no spillover effect of the intervention on parental investments and household responses. The estimation results in table 4 indicate that the intervention had no significant impact on all three variables capturing parental investments and household responses. This insignificant result should be expected since the intervention did not involve the household or parents.

Table 4: The estimation results of the model (3) for both before and after the intervention.

	No. books	No. books	Reading	Reading	Tablet	Tablet
Treat	-0.112	0.152	0.107	0.280	-0.00327	-0.0113
	(0.392)	(0.507)	(0.211)	(0.205)	(0.0425)	(0.0418)
N	688	549	664	530	680	543

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

The next step is to investigate the relative impact of the PDR and the overall quality indices on children's skills. To that end, we estimate the following linear regression.

$$\frac{\theta_{i1}^s - \theta_{i0}^s}{\tau_i} = \alpha^s + \beta_{pdr}^s Q_i^{pdr} + \beta_{overall}^s Q_i^{overall} + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon_i^s, \tag{4}$$

where Q_i^{pdr} and $Q_i^{overall}$ are the PDR and the overall quality indices of student i's classroom and are instrumented by the treatment dummy T_i, all quality indices at the baseline, and evaluator-fixed effects. We consider only the two quality indices because the intervention does not significantly affect the others.

Table 5 presents the estimation results when other quality indices at the endline were not controlled for. The results imply that the plan-do-review activity's quality significantly affects young students' daily learning gains in all three skills. The results also suggest that the impact of the PDR quality on literacy skills is larger than on math skills. It is reasonable because students usually use language to express their thoughts intensively when the plan-do-review (PDR) activity is of high quality. In contrast, the overall quality does not seem to influence the outcomes. Technically, the instruments are sufficiently strong, and the overidentification tests were not rejected.[8]

Similar results (qualitatively at least) are found in table 6, which presents the estimation results when other quality indices at the endline were controlled for. Note that the impact on math skills is not significant for the specification with more controls but still significant for literacy skills. This confirms that the impact of the PDR quality on literacy skills is larger than on math skills, as suggested above. Technically, the overidentification tests were not rejected, but the instruments are weak (F-statistics below 10).

This exploratory analysis suggests that the plan-do-review quality has a stronger impact on children's skills than the overall-classroom quality. This conclusion is only suggestive at best. It is possible that we left

[8] Kleibergen-Paap F-statistics (Kleibergen and Paap, 2006) are well above 10, the conventional threshold proposed by Stock et al. (2002). This implies that the instruments are relevant and strong enough to avoid weak instrument bias. In addition, we cannot reject overidentification tests using Hansen-J statistics for all specifications. This result suggests that our instruments are orthogonal to the error terms.

Table 5: Estimation results for the impacts of PDR and overall quality indices on daily learning gains using the instrumental variable approach.

	COG	COG	MATH	MATH	LIT	LIT
PDR	0.0430***	0.0388***	0.0387***	0.0382***	0.0494**	0.0398*
	(0.0155)	(0.0135)	(0.0142)	(0.0123)	(0.0208)	(0.0207)
Overall	-0.00872	-0.00671	-0.00102	-0.00408	-0.0203	-0.0107
	(0.0126)	(0.0118)	(0.0130)	(0.0127)	(0.0156)	(0.0156)
N	857	547	857	547	857	547
F-Stat.	44.18	23.96	44.18	23.96	44.18	23.96
Over. (p)	0.449	0.418	0.318	0.328	0.187	0.350
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

Table 6: Estimation results for the impacts of PDR and overall quality indices on daily learning gains using the instrumental variable approach with additional controls.

	COG	COG	MATH	MATH	LIT	LIT
PDR	0.0539***	0.0364***	0.0483***	0.0329***	0.0623***	0.0418**
	(0.0120)	(0.0113)	(0.0125)	(0.0111)	(0.0162)	(0.0178)
Overall	0.0175	0.0128	0.0310	0.0273	-0.00269	-0.00882
	(0.0244)	(0.0227)	(0.0258)	(0.0269)	(0.0316)	(0.0290)
N	849	544	849	544	849	544
F-Stat.	5.673	7.436	5.673	7.436	5.673	7.436
Over. (p)	0.583	0.290	0.707	0.541	0.586	0.439
Control	$\tilde{\boldsymbol{X}}^1$	$\tilde{\boldsymbol{X}}^2$	$\tilde{\boldsymbol{X}}^1$	$\tilde{\boldsymbol{X}}^2$	$\tilde{\boldsymbol{X}}^1$	$\tilde{\boldsymbol{X}}^2$

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\tilde{\boldsymbol{X}}^1$ and $\tilde{\boldsymbol{X}}^2$ are the $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ with all three teacher-related quality indices and two school-level quality indices at the endline, where $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

out important learning quality variables, which are also correlated with our excluded instruments, especially the randomized treatment itself. Therefore, this should not be interpreted as conclusive evidence showing that the overall quality has no impact on child development at all. This caution also applies to the other quality dimensions, especially teacher-child interaction and teacher support.

5 Accounting for Concavity of Skill Production Function, and Skill Depreciation

The benchmark specification, model (1), implicitly assumes that the daily learning gain/growth rate is constant with respect to school days or, in other words, the production function of skill is linear in school days. This section extends the model by allowing the production function to be potentially nonlinear. The extension implies that the daily learning gain depends on school days, τ_i. As a result, we add school days as an additional independent variable. The second concern is that the daily learning gain could depend on the child's initial skills. Empirically, we add the baseline skill, θ_{i0}^s as an additional independent variable. This implication can be considered as a convergence of skills related to the convergence implication in the neoclassical growth model.

Another issue is that child skills could be exponentially depreciated over time. More specifically, child i, whose baseline skill was θ_{i0}^s, would have only $e^{-\lambda\tau_i}\theta_{i0}^s$ after τ_i days if she had not been able to learn anything during the time, where λ is the skill depreciation parameter. With this possibility of skill depreciation, the net daily learning gain after τ_i school days can be rewritten as $\frac{\theta_{i1}^s - e^{-\lambda\tau_i}\theta_{i0}^s}{\tau_i}$.

We can now rewrite the benchmark specification using those three implications as follows.

$$\frac{\theta_{i1}^s - e^{-\lambda\tau_i}\theta_{i0}^s}{\tau_i} = \alpha^s + \beta^s T_i + \eta\tau_i + \kappa\theta_{i0}^s + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon_i^s. \tag{5}$$

Using a second-order Taylor approximation, we can express the RHS of the above specification as follows:

$$\frac{\theta_{i1}^s - e^{-\lambda\tau_i}\theta_{i0}^s}{\tau_i} \approx \frac{\theta_{i1}^s - \left(1 - \lambda\tau_i\theta_{i0}^s + \lambda^2\frac{\tau_i^2\theta_{i0}^s}{2}\right)\theta_{i0}^s}{\tau_i} = \frac{\theta_{i1}^s - \theta_{i0}^s}{\tau_i} + \lambda\theta_{i0}^s - \lambda^2\frac{\tau_i\theta_{i0}^s}{2}.$$

As a result, the new specification for the intent-to-treat effect, capturing those three implications, is as follows.

$$\frac{\theta_{i1}^s - \theta_{i0}^s}{\tau_i} = \alpha^s + \beta^s T_i + \eta\tau_i + (\kappa - \lambda)\,\theta_{i0}^s + \lambda^2\frac{\tau_i\theta_{i0}^s}{2} + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon_i^s. \tag{6}$$

Note that the skill depreciation parameter, λ, can be identified using the interaction term of school days and the baseline skill, $\frac{\tau_i\theta_{i0}^s}{2}$, while the convergence parameter κ can be identified using the estimation coefficient of the baseline skill and the estimate of λ. The process would follow a skill convergence process if $\kappa < 0$. In addition, the production function is concave in time if $\eta < 0$.

Table 7 presents the estimation results when we account for the concavity of skill production function, skill convergence, and skill depreciation. First, the negative and significant value of parameter η in all but one specification indicates that the skill production function is concave in time τ_i. See the estimate of η in the second row of the table.

Table 7: Estimation results of model (5) that accounts for the concavity of skill production function, dynamic complementarity, and skill depreciation.

	COG	COG	MATH	MATH	LIT	LIT
Treat	0.0463**	0.0623***	0.0410	0.0578**	0.0480**	0.0614***
	(0.0203)	(0.0195)	(0.0249)	(0.0276)	(0.0199)	(0.0207)
τ_i	-0.00303**	-0.00248**	-0.00266**	-0.00246	-0.00547***	-0.00532***
	(0.00114)	(0.00123)	(0.00131)	(0.00156)	(0.00165)	(0.00138)
θ^s_{i0}	-0.00641***	-0.00605***	-0.00707***	-0.00717***	-0.0109***	-0.0111***
	(0.00115)	(0.00189)	(0.00160)	(0.00215)	(0.00186)	(0.00222)
$\frac{\tau_i \theta^s_{i0}}{2}$	0.0000774***	0.0000615	0.0000721*	0.0000684	0.000151***	0.000154***
	(0.0000273)	(0.0000467)	(0.0000401)	(0.0000537)	(0.0000441)	(0.0000562)
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$
N	866	552	866	552	866	552

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

Second, based on model (5), the skill depreciation parameter, λ, is the square root of the estimated coefficient of the interaction term shown in the last row of the table. The estimate for the basic model of cognitive skills (the last row and the first column) implies that the skill depreciation parameter, $\hat{\lambda} = \sqrt{0.0000774} \approx 0.00880$, which in turn implies that the half-life of the cognitive skills is about $\frac{\ln 2}{\hat{\lambda}} \approx 79$ school days. In other words, half of a child's cognitive skills would be lost if he/she had not been to school for 79 school days (assuming learning nothing outside school). One can speculate further that closing school for a year (roughly 200 school days) would lead to a learning loss in cognitive skills of approximately $\left(1 - e^{-0.00880 \times 200}\right) \times 100 = 82.8\%$, which might be the case for a year-long closure of schools during the COVID-19 pandemic. Another interesting result regarding skill depreciation is that it costs literacy skills the most in both economic and statistical senses. That is, the skill parameter is most prominent in magnitude and most statistically significant for literacy skills. This implies that literacy or language skills in young children depreciate faster than math skills.

We can compute a parameter capturing the convergence process of skill formation using the estimated skill depreciation parameter and the estimation coefficient of the baseline skill. For the basic model of cognitive skills (the first column), the estimate of $\hat{\kappa} = -0.00641 + 0.00880 \approx 0.00159$. The estimate's positivity implies that young children's skill formation exhibits a divergent process. In other words, children with higher initial skills tend to learn faster, i.e., larger daily learning gain.

6 Heterogeneous Effects

This section investigates whether the on-site training benefits the students differently across subgroups, including child gender (Female), student-teacher-ratio (STR), grade levels (Dk3 and Dk3), having special needs (Needs), household wealth (Wealth), main caregiver's education (EduC: dummy for a bachelor degree or above), a fraction of teachers with an ECE degree (ECE), and average teacher age (Tage). Technically, we estimate the heterogeneous effects by adding interaction terms between the treatment and those variables into the benchmark model as follows.

$$\frac{\theta_{i1}^{s} - \theta_{i0}^{s}}{\tau_i} = \alpha^s + \beta^s T_i + \boldsymbol{\phi}^s T_i \times \boldsymbol{H}_i + \boldsymbol{\gamma}^s \boldsymbol{X}_i + \varepsilon_i^s, \tag{7}$$

where $\boldsymbol{H}_i$ is a set of heterogeneous variables of interest. Note that we do not include $\boldsymbol{H}_i$ as an additional set of controls because it is already included in the main set of controls $\boldsymbol{X}_i$. The key parameters of interest, $\boldsymbol{\phi}^s$, estimate the heterogeneous effects of the on-site teacher training on student's skill s.

Table 8 presents the estimation results for the heterogeneous effects (estimated coefficients of the interaction terms). The overall results suggest that the intervention helps students homogeneously, except for a few cases. First, students in the classroom with a larger student-teacher ratio benefit more from the intervention than the others. This effect is significant for literacy skills only. Second, the intervention benefits disadvantaged students (measured by the main caregiver's education) more than others (significant for math only). This suggests that the intervention could be an essential tool for inequality reduction in developing countries. Third, teachers with an ECE degree may learn more from the training and, as a result, can teach young students more effectively.

7 Robustness Checks

This section presents estimation results with changes in several dimensions for robustness checks. The overall results confirm the benchmark results, indicating that the intervention significantly impacts children's skills.

7.1 Estimation Results with Whole Sample

This subsection responds to concerns regarding sample attrition due to missing data. Recall that the original sample contains 998 children, but the main sample has 866 children, or 87 percent of the original, due to missing data. See table A.3 for summary statistics of both samples and the attrition test result.[9] All but three variables are significantly different across samples. In other words, the overall result suggests that the two samples are comparable concerning the observed variables. Nevertheless, we perform the benchmark analysis on the whole sample and present the results in panel A of table 9 below. The results are comparable

[9] We tested for the difference between the two samples by regressing each variable on a dummy variable, indicating whether the observation is in the main sample. The p-value is calculated based on clustering at the school level.

Table 8: Estimation results for heterogeneous effects.

	COG	COG	MATH	MATH	LIT	LIT
TxFemale	-0.00499	0.00242	-0.00654	-0.0154	-0.00267	0.0291
	(0.0238)	(0.0294)	(0.0269)	(0.0330)	(0.0291)	(0.0332)
TxSTR	0.00521	0.00261	0.00240	-0.00116	0.00943***	0.00827**
	(0.00337)	(0.00327)	(0.00378)	(0.00355)	(0.00325)	(0.00344)
TxDk2	-0.0176	-0.0347	-0.0243	-0.0434	-0.00758	-0.0216
	(0.0444)	(0.0510)	(0.0526)	(0.0579)	(0.0443)	(0.0531)
TxDk3	-0.0278	-0.0484	-0.0223	-0.0488	-0.0361	-0.0479
	(0.0340)	(0.0407)	(0.0353)	(0.0439)	(0.0434)	(0.0500)
TxNeeds		-0.00731		-0.00673		-0.00818
		(0.0472)		(0.0549)		(0.0693)
TxWealth		0.0154		0.0260		-0.000552
		(0.0134)		(0.0158)		(0.0171)
TxEduC		-0.128		-0.194*		-0.0298
		(0.0809)		(0.106)		(0.0950)
TxECE		0.0505		0.103**		-0.0288
		(0.0432)		(0.0465)		(0.0570)
TxTage		-0.000252		0.000108		-0.000791
		(0.00134)		(0.00145)		(0.00189)
N	866	552	866	552	866	552
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

to the benchmark results. In this specification, the magnitude of the impact is slightly smaller than the benchmark result in all cases. This small reduction in the impact is likely to result from the fact that the whole sample includes two schools that participated in the on-site training and already implemented the curriculum a few years before the experiment. As a result, there may be no significant change in their classrooms after the intervention.

Table 9: Estimation results for robustness checks.

	COG	COG	MATH	MATH	LIT	LIT
Panel A: ITT with the whole sample						
Treat	0.0519**	0.0532**	0.0407*	0.0435*	0.0688***	0.0677***
	(0.0199)	(0.0212)	(0.0206)	(0.0246)	(0.0241)	(0.0240)
N	998	647	998	647	998	647
Panel B: ITT with IRT Scores						
Treat	0.00126*	0.00239***	0.000907	0.00203**	0.00198**	0.00226**
	(0.000656)	(0.000607)	(0.000654)	(0.000764)	(0.000822)	(0.000859)
N	866	552	866	552	866	552
Panel C: ITT with Endline Scores as the Outcomes						
Treat	3.382**	4.161***	3.384*	4.058**	3.380**	4.315***
	(1.469)	(1.464)	(1.807)	(2.007)	(1.390)	(1.406)
School days	0.0792	0.0885*	0.120**	0.122**	0.0182	0.0377
	(0.0524)	(0.0528)	(0.0572)	(0.0604)	(0.0599)	(0.0678)
Math pre	0.403***	0.379***	0.577***	0.554***	0.142***	0.115***
	(0.0266)	(0.0333)	(0.0315)	(0.0406)	(0.0313)	(0.0376)
Literacy pre	0.370***	0.385***	0.253***	0.273***	0.545***	0.554***
	(0.0313)	(0.0357)	(0.0375)	(0.0452)	(0.0346)	(0.0381)
Panel D: ITT with Total Days						
Treat	0.0236**	0.0299***	0.0182	0.0270**	0.0317**	0.0341**
	(0.0101)	(0.0100)	(0.0111)	(0.0126)	(0.0123)	(0.0130)
N	866	552	866	552	866	552
Control	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$	$\boldsymbol{X}^1$	$\boldsymbol{X}^2$

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001. $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are the same as in table 1.

7.2 Estimation Results when the Test Scores are derived using Item Response Theory (IRT)

The subsection changes the calculation method for the outcome variables. More specifically, the outcomes for cognitive, math, and literacy skills are now derived using item response theory (IRT). This approach allows for different difficulties across items/questions and normalizes the test scores with a unit standard deviation. As a result, an estimated impact is now in a unit of standard deviation. The results in panel B of table 9 are qualitatively similar to those in panel B of table 1. All but one estimate are positive and significant even though their sizes are relatively smaller.

7.3 Estimation Results with Endline Skills as the Outcomes and Baseline Skills and School Days as Additional Controls

The subsection estimates an alternative specification where the outcome is the raw score at the endline and school days, and the raw scores of math and literacy at the baseline are controlled for. See panel C of table 9. Again, the results are qualitatively similar to the benchmark results. A back-of-envelope calculation based on the average school days for the treatment group (67 days) implies that the impact of the intervention of the basic model for cognitive skills in terms of daily learning gains is approximately $\frac{3.387}{67} \approx 0.0505$ per day, which is slightly lower than the benchmark case of 0.0664 per day.

7.4 Estimation Results with the Total Number of Days instead of School Days

The subsection estimates the benchmark model where school days are substituted by the total number of days between the baseline and the endline tests. See panel D of table 9. The results confirm that the intervention significantly impacts children's skills. As expected, the effect size is much smaller in all cases since the learning gain is now divided by a relatively larger number of days.

8 Conclusion

This paper has shown that intensive and hands-on on-site teacher training can effectively improve early childhood education quality and foster young children's cognitive skills. The impact of on-site training for preschool teachers on children's skills is statistically and practically significant. In particular, it caused an increase in the effectiveness of the classroom (measured by daily learning gain) in children's cognitive skills by almost 50 percent relative to the control group. Alternatively, on each day of schooling, treated children accumulated cognitive skills more than the controls by about 0.00302 the control group standard deviations. We also found that the intervention benefits students homogeneously. A general lesson is that the on-site training program is significantly more effective than traditional in-class teacher training, which normally focuses on theoretical and abstract concepts.

Two important on-site training features are (1) its specificity regarding the teaching approach or curriculum and (2) its detailed weekly teaching plans. First, the training primarily focuses on implementing the RIECE curriculum, a HighScope-based curriculum, which mainly focuses on the plan-do-review process (PDR). Even though we could not prove for certain that the positive and significant impact results from the curriculum, we have learned that the intervention significantly improved the plan-do-review (PDR) and the overall-classroom quality, and the former had relatively stronger power to explain the improvement in children's skills. This evidence confirms the important role of the specificity feature of the training in enhancing preschool teachers' skills (see related discussion in Popova et al., 2022). Second, all teachers who attended the on-site training received detailed weekly teaching plans covering all daily activities. These plans represent a complete set of tasks or contents for the teachers to perform daily, which has been proven to be key to success in many education interventions (see for example Andrew et al., 2022; Banerjee et al., 2007). A broad implication is that an effective teacher development program should be specific regarding the teaching approach or curriculum and provide a complete set of tasks or contents consistent with the curriculum.

We also found the main results robust after accounting for the concavity of the skill production function, the convergence of the skill formation process, and skill depreciation. In addition, our findings indicate that the skill formation of young children exhibits a divergent process. That is, children endowed with better initial skills can potentially learn faster (higher daily learning gains). With the divergence of the skill formation process, an effective policy aiming to reduce inequality in human capital should invest early in life (Cunha and Heckman, 2007).

The on-site training intervention is very cost-effective and potentially scalable. The marginal cost related directly to the on-site training was 495 USD per teacher and 32.7 USD per student.[10] This per-student cost is comparable to the cost in Andrew et al. (2022), which was 47 USD per student. On the other hand, it is not straightforward to compare the effectiveness of the two experiments because of their differences in experiment duration. The current intervention was about four months, while the one in Andrew et al. (2022) was longer than one year. To make a reasonable comparison, we have to linearly extrapolate the impact of the current intervention as if it were implemented for one academic year (200 school days). Using the daily learning gain of 0.00302 SD per school day, the estimated total impact of one-year intervention would be about 0.60 SD per academic year, larger than the effect size of 0.17 SD in Andrew et al. (2022). The cost-effectiveness of this experiment should also serve as new evidence showing that the HighScope approach of the Perry Preschool Project (Heckman et al., 2010; Schweinhart and Weikart, 1997) can be effectively

[10]The training fee in 2020 was 15,000 Thai Baht, and the average exchange between Thai Baht and USD (over the year 2020) was 30.29 Thai Baht per USD. Given that the average student-teacher ratio of the treatment group was 15.13, the average cost per student was about 32.7 USD. This excludes costs of other activities, e.g., in-class training, school visits, overhead, and data collection, since the control and treatment share them equally. If all but data collection, overhead, and research costs are included, the cost per head would be at most 75.25 USD per student. This cost is calculated using 998 students in the whole sample.

replicated at scale in rural areas of a developing country. This evidence is more direct to the HighScope approach than the one in Chujan and Kilenthong (2021), where the treatment was to provide a childcare center with an additional teacher trained to implement the RIECE curriculum. The effect size in Chujan and Kilenthong (2021) was roughly 0.40 standard deviation (over eight months), comparable to the current intervention, but it was significantly more costly, with 286 USD per student. Nevertheless, the lesson from the first experiment in Chujan and Kilenthong (2021) was instrumental to the development of the on-site training implemented in the current intervention.

This paper has a couple of limitations. First, it has been shown that improving preschool quality and child development is feasible without affecting school management or teacher incentives. However, this does not imply that those factors are unimportant for student learning outcomes. See, for example, Mbiti et al. (2018). Of course, More research is needed to investigate whether a more comprehensive (more activities) treatment would be more or less cost-effective relative to the current one. On the one hand, it is reasonable to expect a larger impact from such a comprehensive intervention. On the other hand, more activities would come with a higher cost. Second, the two mistakes during the experiment forced us to drop many samples. The timing of the surveys was not ideal either. Measuring young children's skills is an effortful and time-consuming task; therefore, considering the assessment's timing should be researchers' first-order concern. Luckily, the main results still stand regarding both issues. Third, it is difficult to extrapolate the current result to younger children, who, in principle, could potentially benefit more from the intervention. A similar experiment on younger children attending rural childcare centers with more age-appropriate child assessment instruments would be beneficial.

References

Andrew, A., O. Attanasio, R. Bernal, L. C. Sosa, S. Krutikova, and M. Rubio-Codina (2022). Preschool quality and child development. Technical report, National Bureau of Economic Research.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

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

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section 28*(1), 97–104.

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

Chujan, W. and W. T. Kilenthong (2021). Short-term impact of an early childhood education intervention in rural thailand. *Journal of Human Capital 15*(2), 269–290.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015). Early childhood education. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, pp. 235–297. University of Chicago Press.

Heckman, J. J. and D. V. Masterov (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy 29*(3), 446–493.

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.

Kilenthong, W. T., K. Boonsanong, S. Duangchaiyoosook, W. Jantorn, and V. Khruapradit (2023). Learning losses from school closure due to the covid-19 pandemic for thai kindergartners. *Economics of Education Reviews Forthcoming*.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Loyalka, P., A. Popova, G. Li, and Z. Shi (2019). Does teacher training actually work? evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics 11*(3), 128–54.

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2018). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics*.

Popova, A., D. K. Evans, M. E. Breeding, and V. Arancibia (2022). Teacher professional development around the world: The gap between evidence and practice. *The World Bank Research Observer 37*(1), 107–136.

Schweinhart, L. J. and D. P. Weikart (1997). The High/Scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly 12*(2), 117–143.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4), 518–529.

UNICEF (2012). *School Readiness: A Conceptual Framework*. New York: United Nations Children's Fund.

A Additional Tables

Table A.1: Summary statistics of key variables and balanced tests for main sample

	Treatment			Control			Estimation	
	Ave.	SD	N	Ave.	SD	N	Coeff	p-value
From the Baseline								
Cognitive pre	49.822	22.779	338	47.662	21.961	528	2.160	0.494
Math pre	49.507	27.663	338	46.681	25.948	528	2.826	0.435
Literacy pre	50.296	21.016	338	49.134	21.172	528	1.162	0.666
age	5.578	0.650	338	5.510	0.663	528	0.069	0.504
female	0.497	0.501	338	0.491	0.500	528	0.007	0.870
weight	19.321	5.178	338	19.358	5.127	528	-0.037	0.934
height	111.364	6.121	338	110.765	6.722	528	0.599	0.379
sdt ratio	15.130	5.539	27	13.430	6.675	43	1.699	0.293
From the Endline								
special needs	0.070	0.255	315	0.068	0.252	502	0.002	0.943
wealth factor	-0.073	1.135	280	0.001	1.012	468	-0.074	0.454
home urban	0.364	0.482	316	0.257	0.437	526	0.107	0.266
hh size	5.626	1.979	270	5.875	2.000	424	-0.249	0.145
care. edu (college)	0.050	0.219	278	0.052	0.222	445	-0.001	0.948
care. edu (high sch.)	0.201	0.402	278	0.207	0.405	445	-0.005	0.854
Thai at home	0.154	0.362	279	0.131	0.337	467	0.024	0.609
read book	1.948	1.959	259	1.844	1.834	436	0.104	0.626
books at home	2.939	5.417	262	3.025	4.577	426	-0.086	0.837
playing tablet	0.788	0.410	278	0.770	0.422	460	0.018	0.663
Thai at school	0.940	0.237	318	0.865	0.342	489	0.075	0.245
mixed class	0.333	0.480	27	0.419	0.499	43	-0.085	0.527
K2 class	0.222	0.424	27	0.256	0.441	43	-0.034	0.687
K3 class	0.444	0.506	27	0.326	0.474	43	0.119	0.232
teacher age	38.444	11.587	27	41.616	12.154	43	-3.172	0.304
frac. of BA	0.981	0.096	27	0.872	0.310	43	0.109**	0.030
frac. of inclass	0.944	0.212	27	0.779	0.398	43	0.165**	0.021
frac. of ECE	0.759	0.424	27	0.733	0.427	43	0.027	0.798

Note: Clustered-standard errors at the school level are in parentheses: * p<0.05, ** p<0.01,*** p<0.001.

Table A.2: Summary statistics of key variables and balanced tests for whole sample

	Treatment			Control			Estimation	
	Ave.	SD	N	Ave.	SD	N	Coeff	p-value
From the Baseline								
Cognitive pre	48.809	22.378	470	47.662	21.961	528	1.146	0.659
Math pre	48.946	27.061	470	46.681	25.948	528	2.265	0.435
Literacy pre	48.602	21.057	470	49.134	21.172	528	-0.532	0.823
age	5.531	0.649	470	5.510	0.663	528	0.021	0.819
female	0.500	0.501	470	0.491	0.500	528	0.009	0.775
weight	19.427	5.185	470	19.358	5.127	528	0.069	0.862
heightd	111.332	6.164	470	110.765	6.722	528	0.567	0.371
stu-teacher ratio	14.579	5.172	38	13.430	6.675	43	1.149	0.417
From the Endline								
special needs	0.059	0.235	444	0.068	0.252	502	-0.009	0.696
wealth factor	-0.031	1.125	407	0.001	1.012	468	-0.032	0.712
home urban	0.371	0.484	447	0.257	0.437	526	0.115	0.206
hh size	5.583	1.892	386	5.875	2.000	424	-0.292	0.074
care. edu (college)	0.048	0.214	395	0.052	0.222	445	-0.004	0.848
care. edu (high sch)	0.218	0.413	395	0.207	0.405	445	0.011	0.683
Thai at home	0.166	0.373	403	0.131	0.337	467	0.036	0.346
read book	2.138	1.997	378	1.844	1.834	436	0.294	0.140
books at home	3.093	5.404	381	3.025	4.577	426	0.069	0.863
playing tablet	0.792	0.407	403	0.770	0.422	460	0.022	0.512
Thai at school	0.875	0.331	449	0.865	0.342	489	0.010	0.894
mixed class	0.342	0.481	38	0.419	0.499	43	-0.076	0.534
K2 class	0.263	0.446	38	0.256	0.441	43	0.007	0.924
K3 class	0.395	0.495	38	0.326	0.474	43	0.069	0.427
teacher age	38.671	11.603	38	41.616	12.154	43	-2.945	0.287
frac. of BA	0.934	0.207	38	0.872	0.310	43	0.062	0.276
frac. of inclass	0.895	0.288	38	0.779	0.398	43	0.116	0.119
frac. of ECE	0.697	0.458	38	0.733	0.427	43	-0.035	0.713

Note: Clustered-standard errors at the school level are in parentheses: * $p<0.05$, ** $p<0.01$,*** $p<0.001$.

Table A.3: Summary statistics of key variables and attrition test results

	Main Sample			Whole Sample			Estimation	
	Ave.	SD	N	Ave.	SD	N	Coeff	p-value
Cognitive pre	48.505	22.296	866	48.202	22.155	998	2.293	0.546
Math pre	47.784	26.650	866	47.748	26.489	998	0.273	0.948
Literacy pre	49.588	21.107	866	48.883	21.109	998	5.324	0.136
age	5.537	0.658	866	5.520	0.656	998	0.128	0.397
female	0.493	0.500	866	0.495	0.500	998	-0.015	0.631
weight	19.343	5.144	866	19.390	5.152	998	-0.354	0.533
height	110.999	6.497	866	111.032	6.468	998	-0.251	0.791
sdt ratio	16.244	7.545	866	15.943	7.223	998	2.274	0.218
special needs	0.069	0.253	817	0.063	0.244	946	0.038**	0.042
wealth factor	-0.027	1.060	748	-0.014	1.066	875	-0.088	0.450
home urban	0.297	0.457	842	0.309	0.462	973	-0.092	0.527
hh size	5.778	1.994	694	5.736	1.954	810	0.295	0.161
care. edu (college)	0.051	0.221	723	0.050	0.218	840	0.008	0.685
care. edu (high sch.)	0.205	0.404	723	0.212	0.409	840	-0.052*	0.089
Thai at home	0.139	0.347	746	0.147	0.354	870	-0.054	0.220
read book	1.883	1.881	695	1.980	1.916	814	-0.668**	0.010
books at home	2.992	4.910	688	3.057	4.982	807	-0.441	0.503
playing tablet	0.776	0.417	738	0.780	0.415	863	-0.024	0.489
Thai at school	0.895	0.307	807	0.870	0.337	938	0.177	0.214
mixed class	0.345	0.476	866	0.345	0.476	998	0.004	0.979
K2 class	0.225	0.418	866	0.243	0.429	998	-0.138	0.345
K3 class	0.430	0.495	866	0.412	0.492	998	0.134	0.385
teacher age	40.326	11.942	866	40.262	11.964	998	0.489	0.918
frac. of BA	0.912	0.255	866	0.901	0.266	998	0.083	0.427
frac. of inclass	0.831	0.353	866	0.826	0.358	998	0.039	0.762
frac. of ECE	0.747	0.414	866	0.728	0.428	998	0.140	0.348

Note: Clustered-standard errors at the school level are in parentheses: * $p<0.05$, ** $p<0.01$,*** $p<0.001$.