HCEO WORKING PAPER SERIES

Short-term Impact of an Early Childhood Curriculum Intervention in Rural Thailand

Wisuwat Chujan
Weerachart T. Kilenthong

Working Paper

2019-077

12/2019

HUMAN CAPITAL AND ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Short-term Impact of An Early Childhood Curriculum Intervention in Rural Thailand*

Wisuwat Chujan and Weerachart T. Kilenthong

University of the Thai Chamber of Commerce

September 13, 2019

Abstract

This paper evaluates the short-term impact of an early childhood curriculum intervention on child development. Teachers in rural childcare centers in northeastern Thailand were encouraged to employ the new curriculum, which is based primarily on the HighScope approach. We overcome the endogenous decision of teachers to adopt the new curriculum by using the randomization of additional teachers as an instrument. We find that the new curriculum significantly improved child development in several dimensions, including gross motor, fine motor, expressive language, and personal and social skills, with an effect size of roughly 0.54 standard deviations for the benchmark case. The results are robust with regards to various estimation methods,

*We are grateful to all the staff members of the RIECE Thailand project, especially Varunee Kruerpradit, who was instrumental in data collection. We would like to thank Lalita Chanwongpaisarn, Warn N. Lekfuangfu, the editor and anonymous referees for helpful comments. Tee Kilenthong is grateful for financial support from the University of the Thai Chamber of Commerce (UTCC), The Equitable Education Fund (EEF), and Sanguansri and Supachai Suthipongchai. Wisuwat Chujan is grateful for financial support from the Thailand Research Fund through the Royal Golden Jubilee Scholarship Program. The findings and conclusions contained in the report are those of the authors and do not necessarily represent the views of the funders. Email: wisuwat@riped.utcc.ac.th and tee@riped.utcc.ac.th.

child development measures, and sample selections. We also find that the impact of the new curriculum is quite homogeneous across sub-groups except in some dimensions, notably parental absence and teacher's job status.

Keywords: early childhood education; early childhood curriculum intervention; HighScope; child development; developing country; rural development; impact evaluation

JEL classification: I21, J13, J24

1 Introduction

Several early childhood education interventions have been shown to have a positive impact on child development and life-long success (e.g., Campbell et al., 2002; García et al., 2016; Heckman and Masterov, 2007; McKey et al., 1985; Schweinhart et al., 2005; Weikart et al., 1978). One of the most influential projects is the Perry Preschool Project (henceforth, PPP), which is highly cost-effective and therefore has a wide-ranging and long-lasting impact on early childhood education (e.g., Heckman et al., 2010; Schweinhart and Weikart, 1997). Evidence from the project has changed the landscape of education policy around the world. Many countries including Thailand have recently put early childhood education at the forefront of their public policies. However, there are questions as to whether the success of the HighScope Perry Preschool curriculum can be replicated in developing countries or even in developed countries like the US since a recent evaluation of the curriculum using a large dataset from Head Start has found much smaller effects (Walters, 2015).

This paper studies the impact of a HighScope-based curriculum intervention in rural Thailand using the data from the Reducing Inequality through Early Childhood Education program (henceforth, RIECE). The new curriculum, locally called the RIECE curriculum, is based primarily on the HighScope approach, and focused on the Plan-Do-Review learning process (PDR). A key difference between RIECE and the PPP is that the latter complemented the curriculum intervention with weekly home visits while the former did not. The sample size of more than 700 children in this paper is significantly larger than the original sample of 123 children in the PPP (Schweinhart and Weikart, 1997). Our larger sample size should mitigate concerns raised by Charles Murray in Heckman (2013) regarding the small sample size of the original study. However, statistical power in this paper could possibly be lower than the PPP because the randomization in this paper was done at the childcare center level whereas the children were individually randomized in the PPP.[1] To the best of our knowledge, this study is the first multi-site evaluation of a HighScope-based curriculum in a developing country.[2]

The teacher's decision to adopt the new curriculum is endogenous. That is, teachers could choose voluntarily whether and when to adopt the new curriculum. Only 35 percent of classrooms adopted the curriculum, and the earliest adopters started using the curriculum in May while the latest started in December. Note that the curriculum adoption information used in this paper is self-reported by teachers. We overcome the endogenous decision of teachers to adopt the new teaching approach by using the randomization of additional teachers as an instrument. Even though this intervention was not designed as a randomized controlled trial, 19 additional teachers were randomly assigned to 19 childcare centers (out of a total of 50 centers). Importantly, the data confirm that this randomization significantly influences the existing teachers to adopt the new curriculum.

We also analyze the impact of curriculum exposure on child development in Section 4.2. This analysis exploits the fact that, with a relatively large sample, the development outcomes were measured in a period of three months, January to March 2016, and also a large variation of timing of the adoption mentioned earlier. In fact, children's exposure to the new curriculum is varied from one to ten months. This paper should be considered as a short-term impact evaluation because the development outcomes were measured less than a

[1]We in fact address this concern by computing the intraclass correlation of the developmental scores, which are between 0 to 0.053. This indicates that children within classes are not more similar to each other than children in different classes.

[2]The HighScope Preschool Curriculum was implemented in the Eastern Caribbean Area (ECA) on a large scale in 2009-2014 (Nichols et al., 2015), but to date, there has been no impact evaluation on child development outcomes.

year after children were exposed to the intervention. In addition, this paper estimates the heterogeneous effects of the new curriculum. This part investigates whether the benefits of the curriculum are distributed evenly across several subgroups including children characteristics, household characteristics, main caregiver characteristics, mother characteristics, and teacher characteristics.

This paper belongs to limited literature on early childhood education intervention in developing countries (e.g., Andrew et al., 2019; Britto et al., 2017; Malmberg et al., 2011; Moore et al., 2008; Opel et al., 2009; San Francisco et al., 2006). The study that most closely resemble ours is Andrew et al. (2019), which is an evaluation of teacher training, and providing additional resources for materials and new staff in Colombia. The program trains teachers in a treatment group on how to design and implement activities to promote child development. In addition, this paper observed both quantity and quality of teachers' classroom activities. They found that the teacher training improved teaching quality and increased children's cognition, language and school readiness while providing additional resources did not affect child development.

On the other hand, there is a much larger literature on early stimulation interventions in developing countries, which primarily focus on promoting maternal child interaction and improving parenting skills (Attanasio et al., 2018, 2014; Cooper et al., 2009; Gardner et al., 2003; Gertler et al., 2014; Grantham-McGregor et al., 1987, 1991; Powell and Grantham-McGregor, 1989; Walker et al., 2011; Watanabe et al., 2005). Those interventions have been implemented mostly through a home visiting program and usually complemented with a nutritional supplementation. Many of the interventions involved demonstrating play activities to the mother and encouraging her to do the activities with her child. Home-made toys and books were used in addition to items already available at home. The evidence from this literature suggests that early interventions promoting mothers' sensitivity and responsiveness to her infant can have significant benefits on children's mental development and mothers' parenting practices.

The remainder of this paper is organized as follows. Section 2 presents an overview of the intervention and summary statistics. Estimation and identification strategies are discussed in Section 3. Section 4 presents the empirical results, and Section 5, the robustness checks. Section 6 discusses and concludes the paper. All tables are in Appendix A.

2 Background and Data Sources

2.1 Overview of RIECE Thailand

The Reducing Inequality through Early Childhood Education program (RIECE Thailand) aims to improve the quality of early childhood education in rural Thailand. To do so, the project has developed an innovative early childhood curriculum, called the RIECE curriculum, which is based primarily on the HighScope program.[3] The HighScope curriculum aims to support children's cognitive and socioemotional development through active learning where both teachers and children had major roles in shaping children's learning. Children are encouraged to plan, carry out, and reflect on their own activities through a plan-do-review process. Adults observed, supported, and extended children's play through scaffolding. They also encouraged children to make choices, solve problem, and engage in activities. Instead of providing lessons, HighScope emphasizes reflective and open-ended questions asked by teachers. During the first year of implementation, the RIECE curriculum mainly focused on the Plan-Do-Review process (PDR), which is a core activity of HighScope.[4] See Epstein (2012) for details on the HighScope curriculum.

RIECE Thailand was officially launched in May 2015. It covered 50 childcare centers in

[3]See Heckman et al. (2010) and Schweinhart and Weikart (1997) for some key findings regarding the impacts of the HighScope curriculum.

[4]The following are the main differences between the RIECE and HighScope curricular: (1) the RIECE curriculum allows each child to choose only one activity each session while the child is allowed to choose several activities (one activity is also possible) for HighScope; (2) with limited space and a large number of students each class, the RIECE approach limits the number of children in each corner/station while the HighScope has no limit; (3) each child will review verbally in front of the class once a week under the RIECE while each child can choose to review verbally or by drawing with the teacher or in front of the class.

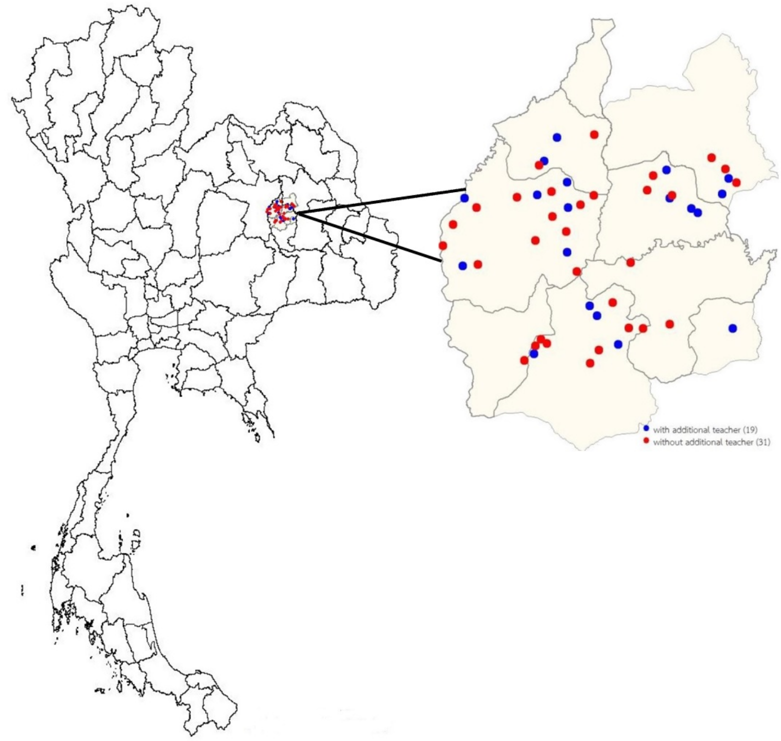


Figure 1: Locations of all participating childcare centers in May 2015.

26 Tambons or subdistricts of Mahasarakham and Kalasin provinces. Figure 1 shows the locations of all participating childcare centers. Most of the centers have two levels of classes, one for 2 to 3 year olds, and the other for 3 to 4 year olds, with only a few exceptions (three centers have four levels up to 6 year olds, and 11 centers have at least one class with mixed-age children). Each center is administered by a subdistrict administration organization (SAO), a local governmental unit in Thailand. Some SAOs administered more than one center.

This project did not randomly assign the new curriculum directly. All existing teachers in all participating centers were encouraged to apply the new curriculum in all classes. All teachers were invited to attend a two-day in-class training (with 98 percent participation rate) and a two-day intensive workshop (with 54 percent participation rate) in April 2015. However, teachers may choose not to adopt the new curriculum; that is, the adoption decision is endogenous. By the end of the academic year, only about 35 percent of classrooms

(45 out of 127) chose to adopt the RIECE curriculum.[5] It is very unlikely that the low adoption is skepticism from parents in this case because Thai rural parents practically have very little involvement with and influence over the center. A potential reason for such a low adoption rate is the lack of effective management by the SAOs. The management issue is very interesting by itself but beyond the scope of this paper. Without an effective management, teachers may choose not to adopt the new curriculum because it generally requires more effort.[6] This self-selection of teachers to adopt the curriculum is the source of the endogeneity problem in this paper.

To overcome this endogeneity problem, we utilize the randomization of additional teachers. Before the first semester of 2015 started, 19 additional teachers[7] were randomly assigned to co-teach in 19 childcare centers (out of 50 participating centers). Importantly, the randomization has been done at the in-class training in April 2015[8] before the semester started without knowing whether each existing teacher will adopt the new curriculum or not. Their main task is to apply the RIECE curriculum in a classroom of 3 to 4 year olds. They also played an important role in transferring knowledge, and supporting and encouraging existing teachers to adopt the new curriculum. The data show that the presence of an additional teacher significantly increases the likelihood of adopting the RIECE curriculum in other classrooms in the same childcare center. Excluding the classrooms with the additional teachers, 42 percent (16 out of 38) of classrooms in the 19 childcare centers with additional teachers adopted the curriculum. In the other 31 centers without additional teachers, the adoption rate was only 14 percent (10 out of 70 classrooms). This randomization of additional teachers

[5] After observing a relatively low adoption rate, the project also invited some existing teachers (no more than four teachers per center) from the centers with no assigned additional teacher to attend an informal five-day on-site training in October that year. We found that more teachers adopted the curriculum after the training.

[6] The intraclass correlation of the adoption decision at the center level is about 0.66 while it is around 0.35 when we consider at the SAO level.

[7] All the additional teachers hold a bachelor's degree in early childhood education from local universities. Most of them graduated in May 2015. They were trained intensively how to apply the RIECE curriculum for two weeks before the semester started.

[8] The Thai school year usually begins in the middle of May and ends in the early of March of next year.

is a potential instrument to overcome the endogeneity problem.

2.2 Child Development Measurement

Child development measures are from the Developmental Surveillance and Promotion Manual (DSPM), developed by the National Institute of Child Health, Department of Health, Ministry of Public Health of Thailand. The DSPM is primarily adapted from the Denver Developmental Screening Test version II or Denver II (Choosri et al., 2017; Morrison et al., 2018). The main purpose of the test is to monitor delayed development in young children.

The DSPM is divided into five main skill domains: gross motor (GM), fine motor (FM), receptive language (RL), expressive language (EL), and personal and social (PS). These domains are found to be key to later academic achievement (Davies et al., 2016). Most of the test items are direct assessments,[9] except some test items in expressive language (EL) and personal and social (PS) domains, which are based on teacher interviews (73 percent of all items for EL and 54 percent of all items for PS). See Appendix A for examples of test items for each domain.

The DSPM is designed for children from birth to 5 years old and categorized into 19 age ranges, each of which may contain several test items for each domain. The details are listed in Appendix A. For each domain, a child is first tested using the test items for his/her own age range. He/She is recorded as passing the test if he/she passes all the test items. Failing at least one item implies that the child has delayed development, and the child is recorded as failing the test.

In order to increase the statistical power, we extended the original DSPM testing procedure by applying not only age-appropriate test items but also test items in two age ranges above or below his/her age range, depending on the test outcomes. If the child passes the test for his/her age range, he/she will be tested using items one level above his/her age

[9]Evaluators or assessors are the interviewers who also collected children and household data. They knew which center received an additional teacher. That is, they were not totally blinded regarding the treatment but they were not supposed to know exactly which classroom has implemented the curriculum.

range. On the other hand, if he/she fails the test for his/her age range, he/she will be tested using items one level below his/her age range. To economize on testing time, we allow only up to two levels above or below an age range. The average testing time was 21 minutes per child with 12.90 standard deviation. The main measure of child development in this paper –the developmental score– is determined by the median of the highest age range that the child passes.

We have transformed the developmental score into an age-standardized score, called the "internally age-standardized score," to deal with differences in the score across ages. Most of the empirical results reported in this paper are based on this standardized score. More formally, let S_{ia} denote the developmental score of child i whose age is in age range a, and $\bar{S}_a$ denote the average score for that age range. The internally age-standardized score for that child is

$$SS_{ia} = \frac{S_{ia} - \bar{S}_a}{\sigma_a} \tag{1}$$

where σ_a denotes the standard deviation of the score for that age range. Note that the standard deviation is the unit of the standardized score.

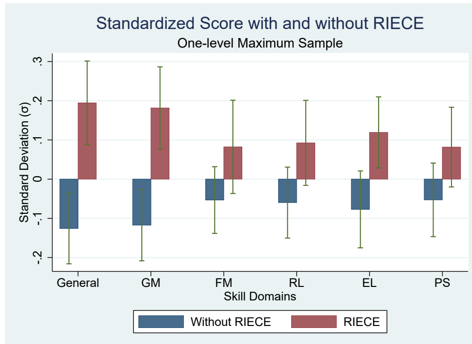
Another important and challenging issue is that the DSPM test is applicable for children up to 60 months old only. Consequently, not all children can be tested up to a maximum of two levels. For example, children older than 55 months can only perform the test for their own age range, and this group consists of roughly 23 percent of the whole sample. These children would generally have lower standardized scores by construction. This, of course, suggests that we should restrict the analysis to the sample of children who can be tested up to a maximum of two levels. However, such a restriction would cost us a significant number of observations. In particular, the whole sample contains 718 children while the two-maximum level sample contains only 323 children.[10] As a compromise, our main results

[10]In principle, one could think of a Tobit-like model to account for the upper bounds of the standardized scores for the whole sample. However, the model would be exceedingly complicated because the upper bounds are varied with child age. As a result, the standard Tobit estimation package is not applicable. Estimating such a complicated model is beyond the scope of this paper.

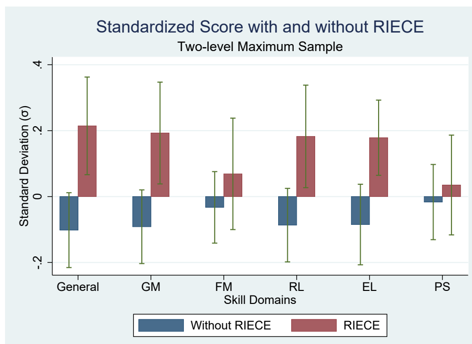
are based on the sample of children who can potentially perform at least one-level above or below their own age range, henceforth called the "one-level maximum sample," which consists of 555 children. Although the one-level maximum sample contains children who can potentially perform more than one level above/below their age range, we calculate the developmental score using up to one-level above/below only. As robustness checks, we also perform main analyses on the other two samples–the "whole sample" and the "two-level maximum sample."

As a first look at the impact of the RIECE curriculum, Figure 2(a) illustrates the standardized scores of children who have learned using the RIECE curriculum and those who have not, using the one-level maximum sample. Note that the "General" variable here represents the average score across all five domains. Figure 2(a) shows that children who were exposed to the RIECE curriculum generally performed better than their peers who were not exposed to the curriculum.[11] Similar patterns emerge for the other two samples–the two-level maximum sample and the whole sample as presented in Figures 2(b) and 2(c), respectively. For each of the five domains, the average standardized score is positive for children who were exposed to the RIECE curriculum, and negative for those who were not exposed to the RIECE curriculum. Note that this information alone cannot answer whether attending regular preschool hampers or benefits child development because this is a relative

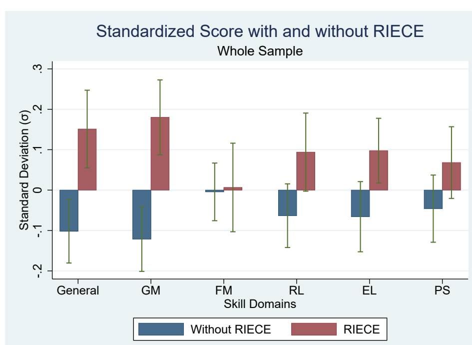
[11] There are at least two key differences between the RIECE curriculum and curricular or practices normally employed in the research area before the intervention. The first and most obvious one is the PDR process, which exists only in the RIECE curriculum. Secondly, under the new curriculum, teachers will systematically plan and regularly implement the daily routine (with several developmentally-stimulating activities each day) starting from early morning until the end of the day almost everyday. Importantly, most of those well-planned activities are designed using active learning concepts, which allows the children to choose materials and decide what to do with the materials. This is clearly different from the standard practice in the area, which resembles the direct instruction concept. More specifically, the majority of the teachers in the area usually will prepare and choose worksheets for children to practice how to write. Some of them may allow children to play with toys but again by choice of teachers not children. Some of the teachers may simply let the children run around the center or watch television almost all day and serving them lunch without well-planned and developmentally-stimulating activities.



(a) One-level maximum sample.



(b) Two-level maximum sample.



(c) Whole sample.

Figure 2: Average standardized scores for children with (red) and without (blue) the RIECE curriculum: (a) using the one-level maximum sample; (b) using the two-level maximum sample; (c) using the whole sample. The error bar represents a 90 percent confidence interval.

score not an absolute one. We can only conclude that children exposed to the RIECE curriculum developed faster than the others. In addition, Table 1 shows that pre-intervention characteristics for both groups of children are not statistically different except the age of the children, parental absence, and main caregiver's education. In other words, the two groups are not significantly different before the introduction of the RIECE curriculum.

2.3 Data on Teachers and the Adoption of the RIECE Curriculum

Data on teachers and the adoption of the RIECE curriculum come from teacher interviews by the survey team of RIECE Thailand. The team began their visits in November 2015, and continued for four rounds until March 2016. The data used in this paper come from the last round of the survey.[12]

One main question in the interview is: "Has your classroom started applying the PDR process in your classroom yet?". If the teacher answers in the affirmative, the teacher would then be asked: "How many days of the week do you apply the curriculum, and in which month did you start using the PDR process?". The data show that among classrooms that adopted the RIECE curriculum, the average number of days per week in which the curriculum is applied is 4.83 days (slightly larger than 4.70 when classrooms with the additional teachers are excluded) and the standard deviation is 0.54. This figure implies that once a teacher decides to use the RIECE curriculum, he/she tends to apply it almost everyday. As a result, this variable is not informative and will be dropped. On the other hand, among classrooms that adopted the RIECE curriculum, children were exposed to the curriculum for an average of 6.5 months and standard deviation of 2.59.[13] There is sufficient variation in the exposure period across classrooms. Therefore, we utilize the number of months as the exposure period. Importantly, the data allow us to match students to teachers and classrooms. By doing so, we can identify whether a sampled child has been exposed to the new curriculum or not. The curriculum adoption dummy variable for each child is our key variable.

[12]The adoption rates of the RIECE curriculum based on other rounds are consistent with the last round.

[13]The number of months is calculated from the first month in which the teacher started using the new curriculum up to the month of the DSPM test for each child.

Another important piece of information is the quality of curriculum adoption. The assessment of the adoption quality was collected by early childhood education experts[14] from RIECE Thailand. The project randomly assigned these experts to visit all 50 centers regularly (on average three times a year). In addition to monitoring and supporting all teachers, these experts assessed the adoption quality as well. Unfortunately, data on the adoption quality is available only for 46 percent of the classrooms. Nevertheless, at the end of year, the expert team did qualitatively evaluate all 50 centers by assigning a score on a discrete scale from zero (worst adoption quality) to one (best adoption quality). The average score of all 50 centers is 0.64 while the minimum and maximum scores are 0 and 1.00, respectively. We calculated the adoption quality score of a particular classroom by multiplying the curriculum adoption dummy variable of that classroom with the adoption quality score of the center to which it belongs.[15]

2.4 Children and Household Data

The key advantage of the RIECE data is that it has information on both schooling and households. The baseline dataset used in this paper is a stratified random sample based on children's age and childcare centers. The data includes no more than 25 randomly selected children from each childcare center. If a center has fewer than 25 children, all children will be selected. Approximately 60 percent of the sample in each center are older than three

[14] There were four early childhood education experts in the team: two of them held a master's degree in early childhood education, and the other two held a bachelor's degree in early childhood education and have a year of experience with the RIECE curriculum.

[15] The weakness of this score is that it represents the quality of the whole center, and not the quality of a specific classroom. Similarly to the adoption decision issue mentioned earlier, the adoption quality of a classroom is potentially different from other classrooms in the same center due to lack of effective management. That is, the adoption quality of the center may not represent the adoption quality of a classroom in that center. Therefore, we only use this (incomplete) piece of information as a robustness check.

years old. The baseline data consist of 1,105[16] children from 1,054 households.

The baseline survey began in April 2015 (before the intervention) and ended in October 2015. The questionnaire comprises three main components: teachers, households, and children. The household questionnaire was developed based on the annual Townsend Thai Data survey. The household questionnaire focuses on socioeconomic status, including household demographics, occupations, labor supply and leisure for each household member, household assets, income, expenditure, borrowing, and lending.

The children questionnaire was developed from several surveys, including the Denver Developmental Screening Test, National Educational Panel Study, World Health Organization Quality of Life, Early Childhood Longitudinal Program, and Cohort Study of Thai Children. For this questionnaire, the respondent must be the main caregiver of the sampled child, who was between 2 and 5 years old at the time. If there is more than one child in a household, the main caregiver will be asked about each child separately. The children questionnaire contains basic information on the children in the household (e.g., age, gender, birth weight, child's health, chronic diseases, disability status, and education attainment), and early childhood investments including time and material inputs, parenting style, and nutritional inputs.

Children's age is measured in months, as of the date of the DSPM test. The low birth weight dummy equals one if the child's birth weight was below 2,500 grams, and zero otherwise (WHO et al., 2006). The sibling dummy equals one if there was at least one sibling living in the same household, and zero otherwise. The dummy for chronic diseases equals one if the child has had asthma, allergies, thalassaemia, glucose-6-phosphate dehydrogenase deficiency (G6PD), anemia, heart disease, epilepsy, tonsillitis, lymphadenitis, pneumonopathy, enteropathy, mycosis, or nephropathy during the last 12 months of the interview, and

[16] Only 735 children (out of 1,105 sampled children) have been tested with the DSPM. The evaluators could not test all the children because the school year ended in early March and it was too complicated to test at their homes. Nevertheless, as shown in Table 1, the summary statistics of the one-level maximum sample (the fourth column) and the whole baseline sample of 1,105 children (the sixth column) are indistinguishable. This should mitigate a concern regarding sample section problem. Further, some key variables are missing for 17 children. As a result, the final whole sample contains 718 children.

zero otherwise. The dummy for vitamin intake is equal to one if the child took a vitamin or supplementary foods during the past 12 months, and zero otherwise. Similarly, each dummy for lego, jigsaw, plastic/wooden shape sorter toy, and clay is defined as one if the household owned at least one piece of the corresponding item. The dummy for parental absence equals to one if the child were living without biological parents at home, and zero otherwise. The dummy for parental marital status equals one if each parent's marital status is divorced, separated, or widowed, and zero otherwise. The dummy for highly-educated mother equals one if the mother finished secondary school, which is the current level of compulsory education in Thailand, and zero otherwise. The dummy for teenage pregnancy mother equals one if the mother was pregnant at age between 10 to 19 years old, and zero otherwise. The dummy for highly-educated main caregiver equals one if the main caregiver finished secondary school, and zero otherwise. Wealth index is constructed by a Principal Component Analysis (PCA) using household asset holdings.[17]

2.5 Summary Statistics of Key Variables

Table 1 presents the summary statistics of key variables related to children categorized by classroom curriculum adoption (with or without the RIECE curriculum). All statistics are calculated from the "one-level maximum" sample of 555 children, which is the sample used in our main analyses. The first column is the treatment group: the children who were exposed to the RIECE curriculum, which is about 40 percent of the sample. The second column is the control group: the children who were not exposed to the curriculum. The summary statistics of the two groups are very similar. The only notable differences between the two groups are the children's age, parental absence, and main caregiver's education. In addition, to answer some concerns regarding the validity of the randomization, the summary statistics of key variables, including children and household variables, and teacher and center characteristics,

[17] Asset holdings used in this exercise include the number of houses/buildings, barns, huts, bicycles, motorcycles, cars, vans/pick-up trucks, motorized carts, farm tractors, four-wheel tractors, trucks, boats, boats with a small motor, telephones, computers/laptops, printers, tablets, air conditioners, cable TVs/satellite dishes, washing machines, televisions, refrigerators, and microwaves.

categorized by the randomization outcome at the center level (centers with or without an additional teacher) are presented in Table 2-4. All tables indicate that the randomization is properly done. All key variables, except low birth weight dummy, sibling dummy, jigsaw dummy, highly educated mother dummy and highly educated main caregiver dummy, are not statistically different.

3 Estimation Methods

This paper estimates the impact of the RIECE curriculum on child development using the following linear model:

$$SS^j_{ia} = \beta_0 + \beta_1 T^j_i + \beta_2 \mathbf{X}^j_i + \varepsilon^j_i \tag{2}$$

where SS^j_{ia} denotes the child development standardized score of child i attending classroom j, $\mathbf{X}^j_i$ denotes a vector of control variables[18] and ε^j_i denotes the error term. The treatment variable of interest is T^j_i, which equals one if child i attending classroom j has been exposed to the new curriculum, and zero otherwise. Our main goal is to identify and estimates β_1, which captures the average treatment effect of the RIECE curriculum on child development.

A statistical challenge in this paper is the endogeneity problem, caused by teachers' decisions to adopt the RIECE curriculum. Unobserved characteristics of teachers, such as their abilities and preferences, can potentially influence their adoption decisions. At the same time, these unobserved characteristics are likely to affect child development independent of their adoption of the curriculum. As a result, the error term ε^j_i, which contains the unob-

[18]The control variables $\mathbf{X}^j_i$ include the student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamin, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, average of teacher's age in each classroom, average of teacher's age in each classroom squared, average of teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

served characteristics, and the curriculum adoption dummy T_i^j are clearly correlated, i.e., $Cov(\varepsilon_i^j, T_i^j) \neq 0$. This correlation leads to an inconsistent estimate of the main parameter of interest, β_1. Intuitively, if a higher-ability teacher is more likely to adopt the new curriculum than a lower-ability teacher, the estimation would be upward-biased. On the contrary, if a low-ability teacher is more likely to adopt, then the estimation would be downward-biased.

We overcome the endogeneity problem by using the randomization of additional teachers as an exclusion restriction. Let RT^j equal one if classroom j belongs to a center that received an additional teacher from the RIECE project, and zero otherwise.

More formally, following Heckman (1976, 1978), we can rewrite the main empirical model together with the adoption decision equation as follows:

$$SS_{ia}^j = \beta_0 + \beta_1 T_i^j + \beta_2 \mathbf{X}_i^j + \varepsilon_i^j \tag{3}$$

where

$$T_i^j = \begin{cases} 1 & if \quad T_i^{j*} > 0 \\ 0 & if \quad T_i^{j*} \leq 0 \end{cases} \tag{4}$$

$$T_i^{j*} = \gamma_0 + \gamma_1 RT^j + \gamma_2 \mathbf{X}_i^j + \boldsymbol{\gamma}_3 \mathbf{Z}^j + \eta_i^j, \tag{5}$$

$\mathbf{Z}^j$ is the vector of control variables[19] for classroom j and η_i^j is the error term. We focus on the estimates of this model using the maximum likelihood estimation (MLE) with normally distributed errors.[20]

For robustness checks, we also estimate model (2) using the instrumental variable approach with the randomization of additional teachers as an instrument. As pointed out

[19]These variables include student-teacher ratio including additional teacher, age level dummies, average of teacher's age in each classroom, average of teacher's age in each classroom squared, average of teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

[20]The joint distribution of the error terms is

$$\begin{bmatrix} \varepsilon \\ \eta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix} \right).$$

by Kelejian (1971), we can consistently identify the main parameter of interest, β_1, using a two-stage least squares (2SLS) approach, in which the first-stage estimation is a simple linear probability model instead of a probit model (see also Heckman 1978). The first-stage regression is the following linear specification:

$$T_i^j = \gamma_0 + \gamma_1 RT^j + \gamma_2 \mathbf{X}_i^j + \gamma_3 \mathbf{Z}^j + \eta_i^j \tag{6}$$

The additional teacher significantly influences the adoption decision of existing teachers in the center (the correlation between RT^j and T_i^j is approximately 0.69). The first-stage estimation results in Table 5 indicate that the instrumental variable is relevant. In particular, the dummy variable of the randomization of additional teachers is statistically significant, and the F-statistic on the excluded instrument is larger than 144, which precludes the possibility of a weak instrument problem.

Moreover, the random assignment feature of additional teachers lends support to the assumption that the assignment of additional teachers is uncorrelated to the error term. Although we cannot verify that groups of children with and without additional teachers are not statistically different in terms of unobservables, we can confirm that almost all observable characteristics of the two groups are not significantly different, as shown in Table 2-4 except low birth weight dummy, sibling dummy, jigsaw dummy, highly educated mother dummy and highly educated main caregiver dummy.

As additional robustness checks, each estimation is also performed on the other two sample sets: (i) the two-level maximum sample, and (ii) the whole sample. In addition, we perform multiple hypothesis correction for p-values using the Holm-Bonferroni method (Holm, 1979).

The instrumental variable approach is also employed to estimate the effects of the length of exposure to the RIECE curriculum. In contrast to the preceding analysis focusing on the extensive margin of adoption, this estimation captures the intensive margin using the number of months as a measure of exposure period. In addition, we estimate the impact of adoption quality on child development using a two-stage least square (2SLS) approach.

We also estimate the effect of the new curriculum on the likelihood of passing the standard age-appropriate DSPM tests. Here, the child development outcome is now a dummy indicating if a child passes his own-age test items for each category. We employ the instrumental variable probit approach (Amemiya, 1978; Newey, 1987).

4 Empirical Results

4.1 The Impacts of the RIECE Curriculum on Child Development: Extensive Margin

This section presents empirical results based on the main empirical model in equations (3)-(5). The results using the one-level maximum sample, shown in Table 6, suggest that the RIECE curriculum has a significant impact on the developmental outcomes of children. The first column indicates that the general score of children using the RIECE curriculum is significantly higher than the non-RIECE group by approximately 0.54 standard deviation.[21] In addition, we find that the RIECE curriculum has positive impacts on child development in all five domains: gross motor (GM), fine motor (FM), receptive language (RL), expressive language (EL), and personal and social (PS). Furthermore, these impacts are statistically significant in all but RL domains. The estimation results with clustered standard errors at both the center and SAO level are presented in Table 7. The standard errors are slightly larger in general but the results are very similar except the insignificance of FM domain.

The 2SLS estimations are very similar to the MLE estimations both in terms of magnitude and significance level. Overall, the results confirm that the RIECE curriculum has a significant impact on child developmental outcomes.

To better understand the endogeneity problem, we present the results from an ordinary least squares (OLS) estimation in Table 6. Qualitatively, the results are quite similar but noticeably different in both magnitude and significance level. In particular, the OLS estimation coefficients for all dimensions are lower than the MLE estimates. This implies that the

[21] A back-of-envelope calculation implies that the RIECE curriculum boosts child development on average by 0.96 months (0.540 multiplied by 1.7793, which is the average standard deviation across all age ranges).

endogeneity problem leads to downward-biased estimates, which are consistent with negative correlations of the error terms ($\rho < 0$) presented in Table 5. This suggests that teachers who adopted the curriculum may have lower (unobserved) ability than the others.

Table 7 presents estimation results when the children in classrooms co-taught by an additional teacher are dropped. This is to answer a concern regarding the fact that an additional teacher may not only affect the adoption decision of existing teachers but also affect child development because he/she also taught the children himself/herself. The results confirmed that the RIECE curriculum significantly raises the developmental outcomes of children. In addition, we also report a reduced form estimation when the curriculum adoption dummy T_i^j is replaced by the randomization outcome RT^j in Table 7. The results again show that the curriculum has a significant but smaller impact on child development relative to the results in Table 6.

4.2 The Impacts of Curriculum Exposure Period on Child Development: Intensive Margin

This section discusses the impacts of curriculum exposure, as measured by the number of months that students have been exposed to the new curriculum. Table 8 shows the results for the one-level maximum sample. The results are qualitatively similar to the results in Table 6. The RIECE curriculum has positive impacts on child development in all five domains. These impacts are statistically significant in all domains except receptive language (RL) and expressive language (EL). More specifically, we find that an additional month of exposure to the RIECE curriculum is correlated with an increase in child development of roughly 0.0626 standard deviation for the general score. To link this intensive margin with the extensive margin discussed earlier, we calculate the average impact by multiplying the intensive margin of 0.0626 standard deviation with the average exposure of 6.35 months,[22] which results in an average impact of about 0.40 standard deviation. This effect is slightly smaller than the

[22]This average exposure is calculated using the exposures of all children who have been exposed to the curriculum.

extensive margin of about 0.47 standard deviation when using 2SLS estimation approach. A potential explanation for the difference is that the relationship between child development and the exposure is non-linear. If it were a linear function, the difference would have been negligible.

4.3 Heterogeneous Effects

This section investigates whether the effects of the new curriculum are heterogeneous across sub-groups. Following the literature, we focus on the heterogeneous effects of children characteristics, household characteristics, main caregiver characteristics, mother characteristics, and teacher characteristics (e.g., Attanasio et al., 2018; Bernal and Keane, 2010; Fiorini, 2010; Fiorini and Keane, 2014; García et al., 2018, 2019; Gregg et al., 2005; Heckman et al., 2010; Kottelenberg and Lehrer, 2014). This section employs the maximum likelihood estimation (MLE) of the following model:

$$SS_{ia}^{j} = \beta_0 + \beta_1 T_i^j + \beta_2 \mathbf{X}_i^j + \beta_3 (T_i^j * H^j) + \beta_4 H^j + \varepsilon_i^j \qquad (7)$$

where

$$T_i^j = \begin{cases} 1 & if \quad T_i^{j*} > 0 \\ 0 & if \quad T_i^{j*} \leq 0 \end{cases} \qquad (8)$$

$$T_i^{j*} = \gamma_0 + \gamma_1 RT^j + \gamma_2 \mathbf{X}_i^j + \boldsymbol{\gamma}_3 \mathbf{Z}^j + \gamma_4 (RT^j * H^j) + \gamma_5 H^j + \eta_i^j, \qquad (9)$$

H^j is a variable of interest capturing the heterogeneity, which includes boy dummy, low birth weight dummy, chronic disease dummy, household wealth index, parental absence dummy, parental marital status dummy, main caregiver's education, main caregiver's age, digit-span recall score of main caregiver, highly educated mother dummy, mother's age, teenage pregnancy mother dummy, average teacher's experience, fraction of tenured teachers, and fraction of teachers with bachelor degrees in early childhood education, and $\mathbf{X}_i^j$ and $\mathbf{Z}^j$ are the vectors of control variables, specified earlier. See also Angrist and Pischke (2008); Blundell et al. (2005) for an instrumental variable approach with interaction terms. The coefficient of the interaction term, (β_3), will be interpreted as the measure of the heterogeneous effect for the variable of interest. Note that these heterogeneity variables are considered each one at a time.

The estimation results, presented in Table 9, show that the estimation coefficients for the interaction terms are not statistically significant in most cases. This implies that the impact of the new curriculum is quite homogeneous across sub-groups except in some domains of the following eight dimensions: child gender, parental absence, main caregiver's digit-span memory score, mother's education, mother's age, teenage pregnancy status of mother, average teacher's experience and fraction of tenured teachers. Child gender is one of the most widely analyzed variable for heterogeneous effects in the education literature (e.g., Attanasio et al., 2018; Fiorini and Keane, 2014; García et al., 2018). The estimation results in the first panel of the table show that boys and girls benefit from the intervention equally in all but PS domain. The estimation coefficient of the interaction term with the boy dummy for PS domain is positive and statistically significant. The positive and significant result is consistent with García et al. (2018) who show that boys benefit relatively more from a high-quality center-based childcare program using the data from ABC/CARE program.

Table 9 also shows that the estimation coefficients of the interaction terms with the parental absence dummy are positive and statistically significant for the general score, EL, and PS domains. In words, the introduction of the new curriculum is beneficial to children whose biological parents live outside the household.[23] A potential reason is that these left-behind children may lack active and positive interaction at home (relative to the other group) because grandparents, who are their main caregivers, might be too old[24] to perform child-appropriate activities or actively respond to child-learning demand. This lack of opportunity could then lead to a delay in child development. Therefore, the introduction of the new curriculum with active and positive interaction at school may benefit these disadvan-

[23]Approximately 45 percent of children in the survey area were living without biological parents at home (Dinh and Kilenthong, 2018). This is an important issue for developing countries but has not been a major focus in the literature.

[24]The average age of main caregivers for the children living without biological parents at home is about 53 years old while the other group is around 36 years old (Dinh and Kilenthong, 2018).

tage children more than the others because they started lower, as suggested in Cunha and Heckman (2009). This implies that the curriculum intervention here can reduce the child development gap between children whose parents live within and outside the household. The same explanation can help explain the negative effect of the interaction term with the main caregiver's digit-span memory score because older people tend to have a lower digit-span memory score.

Another interesting issue is about mother characteristics including mother's education, mother's age, teenage pregnancy status of mother.[25] In this case, we focus only on children whose mothers live in the households (73 percent are the main caregiver). This is more reasonable than working with all samples because mothers who do not live in the households would not be able to perform child-appropriate activities with the children, and therefore, should not have a significant impact on child development. Table 9 shows that the estimation coefficients of the interaction terms with mother's education, mother's age, teenage pregnancy status of mother are statistically significant for FM; general, GM and FM; and FM and EL domains, respectively. In fact, all three variables are capturing mothers' ability to raise a child, and therefore, largely correlated. That is, mothers with higher education are more likely to be older and less likely to be teenage pregnancy mothers. This explains why the coefficients for the interaction terms with mother's education and mother's age are positive while teenage pregnancy status of mother is negative. Together, the results imply that the introduction of the new curriculum and mothers' ability to raise a child are complementary.

Panel (5) of Table 9 reports that the estimation coefficients of the interaction terms with fraction of tenured teachers are positive and statistically significant for general, FM and EL domains. A potential reason is that tenured teachers may have higher (unobserved) teaching ability, and therefore, may be able to implement the new curriculum more effectively. This result is consistent with the heterogeneous effect of teacher's experience, which is positive and statistically significant for FM domain. We must be careful in drawing a policy impli-

[25] A teenage pregnancy mother is a mother who was pregnant at age between 10 to 19 years old.

cation from these results, however. Even though they may have high potentials to generate good outcomes, the tenure status may reduce their incentives to adopt an effective teaching approach that requires more effort. In fact, our data confirm that tenured teachers are less likely to adopt the new curriculum.

5 Robustness Checks

5.1 The Impacts of Adoption Quality

Another variable of interest is the quality of curriculum adoption. Due to data limitation, we generate a measure of adoption quality by interacting the center's adoption quality score with the classroom's curriculum adoption dummy. That is, only classrooms that adopted the RIECE curriculum will have a positive adoption quality score while the others will have a zero score. We should be careful of interpreting this quality variable, however. As mentioned earlier, the adoption quality is quite heterogeneous across classrooms in the same center. As a result, the adoption quality of the center may not represent the adoption quality of each classroom in that center. Nevertheless, it should at least contain some information reflecting the true quality. We expect adoption quality to have a significant effect on child development.

Table 10 shows the results for the one-level maximum sample from a two-stage least squares (2SLS) estimation. Adoption quality significantly improves child development. In particular, a one-percentage-point increase in adoption quality significantly boosts overall child developmental score by 0.0059 standard deviation (the adoption quality used in the estimation ranged from 0 to 1). As before, adoption quality has positive effects on developmental scores in all five domains, and these effects are statistically significant in all domains except receptive language (RL). As in the case of the intensive margin, we can calculate the average impact by multiplying the quality effect of 0.59 standard deviation with the average adoption quality of 0.86, which results in an average impact of about 0.51 standard deviation. This effect is slightly lower than the extensive margin of about 0.47 standard deviation when using 2SLS estimation approach. Again, a potential explanation for the difference is that the relationship between child development and the adoption quality is non-linear.

5.2 Alternative Samples

This section replicates the main analyses using the two-level maximum and whole samples. Table 6 shows that the RIECE curriculum has significant effects on child development in the general, GM, FM and EL domains for these two alternative samples. On the other hand, the effect of the curriculum on FM and PS domains are statistically significant only for the two-level maximum and the whole samples, respectively. Nevertheless, the estimates in all five domains have positive signs even though some are not statistically significant.

Similarly, the impacts of curriculum exposure and adoption quality on child development are qualitatively comparable to the benchmark cases with the one-level maximum sample. The results for curriculum exposure are shown in Table 8, and the results for adoption quality are shown in Table 10.

To summarize, we find that the RIECE curriculum significantly improves several dimensions of child development using different samples.

5.3 The Original Measure of Child Development for DSPM: Pass/Fail Outcomes

We estimate the effects of the RIECE curriculum on the likelihood of passing the standard age-appropriate tests as another robustness check. Note that the overall development variable, "General" in this case equals one if the child passes the standard age-appropriate tests for all five domains and zero otherwise. We employ an instrumental variable probit model (Amemiya, 1978; Newey, 1987). The child development outcome for each domain is now a dummy variable that equals one if the child passes his own-age items in the domain, and zero otherwise.

Table 11 presents the marginal effects (at the mean) of the new curriculum on the likelihood of passing the test. The results are based on the one-level maximum sample. The impact on PS domain is positive and statistically significant while the rest are positive but insignificant. A potential explanation for these insignificance results is that the standard DSPM test by itself has low statistical power by design, as discussed earlier in Section 2.2.

6 Discussion and Conclusion

This paper has shown that the RIECE curriculum, a HighScope-based curriculum, significantly improved multiple dimensions of child development, including gross motor, fine motor, expressive language, and personal and social skills. The results are robust to various estimation methods, child development measures, and sample selections. In addition, we found that the impact of the new curriculum is quite homogeneous across sub-groups except in some dimensions, notably parental absence and teacher's job status.

These findings complement the findings of the Perry Preschool Project (Heckman et al., 2010; Schweinhart and Weikart, 1997) but in the context of a developing country. In other words, this paper has provided evidence that the success of the HighScope Perry Preschool project can be replicated in a developing country with a relatively large effect size of approximately 0.54 standard deviation. On the other hand, the intervention cost in the first year is at about 243 USD per year per (treated) child[26] (counting only children who have been exposed to the curriculum during the first year). This cost is relatively low compared to comparable interventions such as Attanasio et al. (2014) which costs around 500 USD per year per child with an effect size of 0.26 standard deviation and Attanasio et al. (2018) which costs around 339 USD per year per child with an effect size of 0.15 standard deviation. From the cost-effectiveness perspective, this curriculum intervention is highly effective and financially feasible for scaling up in developing countries. However, the problem of low adop-

[26] The total cost of the intervention (excluding data collection cost) for the first year of the project was approximately 7.9 millions Baht, which can be decomposed into 5.5 millions Baht for personnel cost, and 2.4 millions Baht for operation cost. The total number of children who were exposed to the new curriculum was 935 children (out of 2,696 children attended all 50 centers). Therefore, the average cost of the intervention for the first year was approximately 8,450 Baht per (treated) child, which is equivalent to 243 USD per child (the average exchange rate during the year 2015 and 2016 was 34.77 Baht per USD).

tion rate has to be resolved. Perhaps, more resources may be needed to boost the adoption rate, perhaps through more effective training, better incentive systems, or more effective monitoring systems etc.

One key limitation of this paper is the incomplete data on the quality of curriculum adoption. Ideally, we would have data on adoption quality for each classroom, and adoption quality would be measured along several dimensions. Unfortunately, more detailed investigations on the implementation of the curriculum are beyond the scope of this paper.

Another limitation is related to the child development measurement itself. The DSPM test is just one of many available tests, e.g., executive functions, behavioral problem index, cognitive skills and non-cognitive skills. With more resources, the RIECE project should conduct different tests to establish whether our findings are sensitive to development measures. Moreover, some of the tests, e.g., Mathematics, Sciences, and Language, are more appropriate for older children. The RIECE project should apply these tests to ascertain the impact of the curriculum on elementary school or middle school children.

References

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. *Econometrica: Journal of the Econometric Society*, pages 1193–1205.

Andrew, A., Attanasio, O., Bernal, R., Sosa, L. C., Krutikova, S., and Rubio-Codina, M. (2019). Preschool Quality and Child Development. Technical report, National Bureau of Economic Research.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., and Rubio-Codina, M. (2018). Early Stimulation and Nutrition: The Impacts of a Scalable Intervention. Technical report, National Bureau of Economic Research.

Attanasio, O. P., Fernández, C., Fitzsimons, E. O., Grantham-McGregor, S. M., Meghir, C., and Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *Bmj*, 349:g5785.

Bernal, R. and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1):164–189.

Blundell, R., Dearden, L., and Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3):473–512.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C., et al. (2017). Nurturing care: promoting early childhood development. *The Lancet*, 389(10064):91–102.

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

Choosri, N., Pookao, C., Swangtrakul, N., and Atkin, A. (2017). Tangible interface game for stimulating child language cognitive skill. *IADIS International Journal on WWW/Internet*, 15(2).

Cooper, P. J., Tomlinson, M., Swartz, L., Landman, M., Molteno, C., Stein, A., McPherson, K., and Murray, L. (2009). Improving quality of mother-infant relationship and infant attachment in socioeconomically deprived community in South Africa: randomised controlled trial. *Bmj*, 338:b974.

Cunha, F. and Heckman, J. J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association*, 7(2-3):320–364.

Davies, S., Janus, M., Duku, E., and Gaskin, A. (2016). Using the Early Development Instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly*, 35:63–75.

Dinh, N. T. T. and Kilenthong, W. T. (2018). Do parental absence and children's gender affect early childhood investment? Evidence from rural Thailand. Working paper.

Epstein, A. S. (2012). *The HighScope Preschool Curriculum, Creative Arts*. HighScope Press.

Fiorini, M. (2010). The effect of home computer use on children's cognitive and non-cognitive skills. *Economics of Education review*, 29(1):55–72.

Fiorini, M. and Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, 32(4):787–836.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2016). The life-cycle benefits of an influential early childhood program. Technical report, National Bureau of Economic Research.

García, J. L., Heckman, J. J., and Ziff, A. L. (2018). Gender differences in the benefits of an influential early childhood program. *European economic review*, 109:9–22.

García, J. L., Heckman, J. J., and Ziff, A. L. (2019). Early childhood education and crime. *Infant mental health journal*, 40(1):141–151.

Gardner, J. M., Walker, S. P., Powell, C. A., and Grantham-McGregor, S. (2003). A randomized controlled trial of a home-visiting intervention on cognition and behavior in term low birth weight infants. *The Journal of pediatrics*, 143(5):634–639.

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

Grantham-McGregor, S., Schofield, W., and Powell, C. (1987). Development of severely malnourished children who received psychosocial stimulation: six-year follow-up. *Pediatrics*, 79(2):247–254.

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

Gregg, P., Washbrook, E., Propper, C., and Burgess, S. (2005). The effects of a mother's return to work decision on child development in the UK. *The Economic Journal*, 115(501):F48–F80.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4):931–959.

Heckman, J. J. (2013). *Giving kids a fair chance*. Mit Press.

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

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

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *Journal of the American Statistical Association*, 66(334):373–374.

Kottelenberg, M. J. and Lehrer, S. F. (2014). The gender effects of universal child care in Canada: Much ado about boys. *Unpublished manuscript, Department of Economics, Queens University.*

Malmberg, L.-E., Mwaura, P., and Sylva, K. (2011). Effects of a preschool intervention on cognitive development among East-African preschool children: A flexibly time-coded growth model. *Early Childhood Research Quarterly*, 26(1):124–133.

McKey, R. H. et al. (1985). The impact of Head Start on children, families and communities. Final report of the Head Start evaluation, synthesis and utilization project.

Moore, A. C., Akhter, S., and Aboud, F. E. (2008). Evaluating an improved quality preschool program in rural Bangladesh. *International Journal of Educational Development*, 28(2):118–131.

Morrison, J., Chunsuwan, I., Bunnag, P., Gronholm, P. C., and Estrin, G. L. (2018). Thailand's national universal developmental screening programme for young children: action research for improved follow-up. *BMJ global health*, 3(1):e000589.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36(3):231–250.

Nichols, T., Greene, J., and Boeckmann, S. (2015). Final Report For the Formative Evaluation of the High/Scope Curriculum Reform Programme. Technical report, UNICEF Office for the Eastern Caribbean Area, BARBADOS.

Opel, A., Ameer, S. S., and Aboud, F. E. (2009). The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children. *International Journal of Educational Research*, 48(1):12–20.

Powell, C. and Grantham-McGregor, S. (1989). Home visiting of varying frequency and child development. *Pediatrics*, 84(1):157–164.

San Francisco, A. R., Arias, M., Villers, R., and Snow, C. (2006). Evaluating the impact of different early literacy interventions on low-income Costa Rican kindergarteners. *International Journal of Educational Research*, 45(3):188–201.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., and Nores, M. (2005). Lifetime effects: The High/Scope Perry preschool study through age 40 (monographs of the High/Scope educational research foundation, 14). *Ypsilanti, MI: High Scope Educational Research Foundation*.

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

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., et al. (2011). Inequality in early childhood: risk and protective factors for early child development. *The Lancet*, 378(9799):1325–1338.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. *American Economic Journal: Applied Economics*, 7(4):76–102.

Watanabe, K., Flores, R., Fujiwara, J., and Tran, L. T. H. (2005). Early childhood development interventions and cognitive development of young children in rural Vietnam. *The Journal of nutrition*, 135(8):1918–1925.

Weikart, D. P., Bond, J. T., and McNeil, J. T. (1978). *The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade.* Number 3. High/Scope Foundation.

WHO et al. (2006). *WHO child growth standards: length/height for age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age, methods and development.* World Health Organization.

A Tables

Table 1: Summary Statistics of Key Variables Categorized by Curriculum

VARIABLES	With RIECE	Without RIECE	Total	Number of Observations	Baseline 2015	Baseline 2015 Observations
Boy dummy	0.48	0.50	0.49	555	0.51	1,105
	(0.50)	(0.50)	(0.50)		(0.50)	
Child's age (months)	47.03	45.53**	46.12	555	48.06	1,061
	(5.91)	(5.82)	(5.89)		(8.32)	
Low birth weight dummy	0.10	0.07	0.08	555	0.09	1,081
	(0.30)	(0.26)	(0.28)		(0.29)	
Sibling dummy	0.43	0.48	0.46	555	0.47	1,105
	(0.50)	(0.50)	(0.50)		(0.50)	
Chronic disease dummy	0.12	0.10	0.11	555	0.12	1,105
	(0.32)	(0.31)	(0.31)		(0.32)	
Vitamin intake dummy	0.55	0.61	0.58	555	0.57	1,104
	(0.50)	(0.49)	(0.49)		(0.49)	
Lego dummy	0.58	0.56	0.57	555	0.56	1,105
	(0.50)	(0.50)	(0.50)		(0.50)	
Jigsaw dummy	0.25	0.29	0.27	555	0.28	1,104
	(0.43)	(0.45)	(0.45)		(0.45)	
Shape dummy	0.26	0.29	0.28	555	0.27	1,105
	(0.44)	(0.46)	(0.45)		(0.45)	
Clay dummy	0.59	0.55	0.57	555	0.59	1,105
	(0.49)	(0.50)	(0.50)		(0.49)	
Parental absence dummy	0.39	0.47†	0.44	554	0.44	1,103
	(0.49)	(0.50)	(0.50)		(0.50)	
Divorced, separated, or widowed parent dummy	0.31	0.28	0.29	553	0.30	1,103
	(0.46)	(0.45)	(0.45)		(0.46)	
Mother's highly educated dummy	0.57	0.50	0.53	503	0.57	1,055
	(0.50)	(0.50)	(0.50)		(0.50)	
Mother's age	29.27	29.11	29.18	450	29.46	956
	(6.27)	(5.88)	(6.04)		(6.09)	
Teenage pregnancy mother dummy	0.25	0.21	0.22	450	0.21	923
	(0.43)	(0.41)	(0.42)		(0.41)	
MC's highly educated dummy	0.31	0.23*	0.26	498	0.26	1,076
	(0.47)	(0.42)	(0.44)		(0.44)	
MC's age	43.34	43.86	43.66	495	44.32	1,071
	(13.72)	(12.67)	(13.07)		(13.29)	
MC's digit-span recall score	3.21	3.28	3.25	355	3.21	762
	(1.46)	(1.58)	(1.54)		(1.51)	
Wealth index	0.07	-0.04	0.00	507	-0.00004	1,104
	(0.90)	(0.88)	(0.89)		(0.89)	
Fraction of sample	39.28%	60.72%	100%			

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, † $p<0.1$. Standard deviations are in parentheses. The first two columns report the statistics for children who exposed to the RIECE curriculum and children who did not, respectively. The third column shows the statistics for all children in the one-level maximum sample. The fifth and the sixth column show the statistics and the number of observations for all children in the baseline 2015, respectively.

Table 2: Summary Statistics of Key Variables Categorized by Additional Teacher

VARIABLES	Centers with Additional Teacher	Centers without Additional Teacher	Total	Number of Observations
Boy dummy	0.50	0.49	0.49	555
	(0.50)	(0.50)	(0.50)	
Child's age (months)	46.27	46.00	46.12	555
	(6.02)	(5.80)	(5.89)	
Low birth weight dummy	0.11	0.06†	0.08	555
	(0.31)	(0.25)	(0.28)	
Sibling dummy	0.42	0.49†	0.46	555
	(0.49)	(0.50)	(0.50)	
Chronic disease dummy	0.10	0.12	0.11	555
	(0.30)	(0.32)	(0.31)	
Vitamin intake dummy	0.56	0.60	0.58	555
	(0.50)	(0.49)	(0.49)	
Lego dummy	0.53	0.59	0.57	555
	(0.50)	(0.49)	(0.50)	
Jigsaw dummy	0.22	0.32**	0.27	555
	(0.41)	(0.47)	(0.45)	
Shape dummy	0.27	0.29	0.28	555
	(0.44)	(0.45)	(0.45)	
Clay dummy	0.58	0.55	0.57	555
	(0.49)	(0.50)	(0.50)	
Parental absence dummy	0.42	0.45	0.44	554
	(0.49)	(0.50)	(0.50)	
Divorced, separated, or widowed parent dummy	0.31	0.28	0.29	553
	(0.46)	(0.45)	(0.45)	
Mother's highly educated dummy	0.59	0.48*	0.53	503
	(0.49)	(0.50)	(0.50)	
Mother's age	29.32	29.06	29.18	450
	(6.09)	(6.00)	(6.04)	
Teenage pregnancy mother dummy	0.24	0.21	0.22	450
	(0.43)	(0.41)	(0.42)	
MC's highly educated dummy	0.30	0.23†	0.26	498
	(0.46)	(0.42)	(0.44)	
MC's age	44.18	43.26	43.66	495
	(12.97)	(13.15)	(13.07)	
MC's digit-span recall score	3.31	3.21	3.25	355
	(1.51)	(1.56)	(1.54)	
Wealth index	0.04	-0.03	0.00	507
	(0.96)	(0.84)	(0.89)	
Fraction of sample	44.14%	55.86%	100%	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Standard deviations are in parentheses. The first two columns report the statistics for children in childcare centers that received an additional teacher and the ones that did not, respectively. The third column shows the statistics for the one-level maximum sample.

Table 3: Summary Statistics of Key Variables for Teacher

VARIABLES	Centers with Ramdomized Teacher	Centers without Ramdomized Teacher	Total	Number of Observations
Average teacher's age (years)	41.93	42.19	42.07	191
	(8.75)	(8.22)	(8.43)	
Fraction of female teachers	0.96	0.95	0.95	193
	(0.19)	(0.23)	(0.21)	
Average teacher experience in early	12.66	13.79	13.29	190
childhood education (years)	(7.18)	(8.06)	(7.69)	
Fraction of tenured teachers	0.49	0.53	0.51	190
	(0.50)	(0.50)	(0.50)	
Fraction of teachers with	0.81	0.73	0.76	191
bachelor degrees in early childhood	(0.40)	(0.45)	(0.43)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Standard deviations are in parentheses. The first two columns report the statistics for teachers in childcare centers that received an additional teacher and the ones that did not, respectively. The third and the fourth columns show the statistics and the number of teachers excluding additional teachers, respectively.

Table 4: Summary Statistics of Key Variables for Childcare Center

VARIABLES	Centers with Ramdomized Teacher	Centers without Ramdomized Teacher	Total	Number of Observations
Fraction of classrooms with	0.07	0.10	0.09	127
mixed-age children	(0.26)	(0.30)	(0.28)	
Fraction of classrooms for	0.37	0.39	0.38	127
3 to 4 year olds children	(0.49)	(0.49)	(0.49)	
Average number of classrooms	3.00	2.26	2.54	50
in a center	(2.58)	(1.00)	(1.79)	
Student-teacher ratio	13.42	14.87	14.32	50
(excluding additional teachers)	(3.76)	(3.65)	(3.72)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Standard deviations are in parentheses. The first two columns report the statistics for classrooms in childcare centers that received an additional teacher and the ones that did not, respectively. The third and the fourth columns show the statistics and the number of corresponding observations. The average student-teacher ratio is calculated based on student-teacher ratio in each center excluding the additional teachers.

Table 5: First-Stage Estimation (One-level Maximum Sample)

	Maximum Likelihood Estimation						2SLS
VARIABLES	General	GM	FM	RL	EL	PS	
Additional Teachers	2.934***	2.886***	2.881***	2.892***	2.889***	2.914***	0.6898***
	(0.236)	(0.235)	(0.236)	(0.232)	(0.232)	(0.235)	(0.0395)
Child's age	0.0719	0.0816	0.0985	0.0729	0.0741	0.0762	0.0100
	(0.167)	(0.173)	(0.165)	(0.172)	(0.173)	(0.172)	(0.0339)
Child's age squared	-0.000923	-0.00105	-0.00128	-0.000963	-0.000962	-0.000965	-0.0001
	(0.00189)	(0.00195)	(0.00187)	(0.00194)	(0.00195)	(0.00194)	(0.0004)
Boy dummy	-0.107	-0.106	-0.0852	-0.105	-0.113	-0.129	-0.0216
	(0.149)	(0.150)	(0.149)	(0.150)	(0.150)	(0.150)	(0.0283)
Low birth weight dummy	-0.0726	-0.0818	-0.0842	-0.0902	-0.0995	-0.0716	-0.0287
	(0.248)	(0.252)	(0.257)	(0.254)	(0.255)	(0.245)	(0.0560)
Sibling dummy	-0.243	-0.235	-0.226	-0.234	-0.235	-0.278†	-0.0179
	(0.152)	(0.153)	(0.152)	(0.153)	(0.153)	(0.153)	(0.0282)
Chronic disease dummy	0.631**	0.593*	0.584*	0.596*	0.603*	0.666**	0.0756
	(0.244)	(0.246)	(0.246)	(0.247)	(0.247)	(0.248)	(0.0475)
Vitamin intake dummy	-0.161	-0.156	-0.156	-0.151	-0.160	-0.151	-0.0340
	(0.155)	(0.156)	(0.156)	(0.156)	(0.156)	(0.154)	(0.0300)
Lego dummy	0.264	0.254	0.258	0.249	0.267	0.250	0.0572†
	(0.172)	(0.172)	(0.172)	(0.171)	(0.174)	(0.169)	(0.0318)
Jigsaw dummy	0.0459	0.0323	0.0454	0.0204	0.0208	0.00366	-0.0035
	(0.189)	(0.195)	(0.191)	(0.192)	(0.192)	(0.184)	(0.0342)
Shape dummy	-0.279	-0.267	-0.282	-0.253	-0.266	-0.285	-0.0372
	(0.186)	(0.187)	(0.187)	(0.187)	(0.186)	(0.187)	(0.0344)
Clay dummy	-0.0417	-0.0435	-0.0288	-0.0414	-0.0433	-0.0104	-0.0059
	(0.150)	(0.152)	(0.151)	(0.152)	(0.152)	(0.153)	(0.0293)
Class type							
3-4 year-olds	1.511***	1.548***	1.565***	1.567***	1.541***	1.542***	0.2621***
	(0.239)	(0.237)	(0.228)	(0.238)	(0.239)	(0.233)	(0.0399)
Mixed	1.189***	1.265***	1.189***	1.257***	1.257***	1.219***	0.2120***
	(0.217)	(0.210)	(0.210)	(0.213)	(0.211)	(0.213)	(0.0357)
4-5 year-olds	0.913†	0.997†	0.995*	1.008†	0.986†	0.807	0.1372†
	(0.515)	(0.511)	(0.496)	(0.516)	(0.515)	(0.523)	(0.0814)
Student-teacher ratio	0.0729***	0.0686***	0.0693***	0.0683***	0.0683***	0.0738***	0.0106*
	(0.0207)	(0.0206)	(0.0210)	(0.0203)	(0.0203)	(0.0204)	(0.0043)
Average teacher's experience	0.0545*	0.0575**	0.0556**	0.0572**	0.0574**	0.0593**	0.0045
in classroom	(0.0217)	(0.0221)	(0.0211)	(0.0217)	(0.0218)	(0.0214)	(0.0044)
Average teacher's age	-0.340**	-0.334**	-0.351**	-0.333**	-0.348**	-0.342**	-0.0436*
in classroom	(0.114)	(0.116)	(0.113)	(0.116)	(0.115)	(0.114)	(0.0200)
Average of teacher's age in each	0.00358*	0.00350*	0.00373**	0.00349*	0.00365*	0.00357*	0.0005†
classroom squared	(0.00145)	(0.00145)	(0.00143)	(0.00146)	(0.00145)	(0.00143)	(0.0003)
Fraction of teacher who	1.257***	1.259***	1.262***	1.241***	1.253***	1.250***	0.1890***
finished early childhood education	(0.307)	(0.315)	(0.313)	(0.316)	(0.312)	(0.296)	(0.0576)
Fraction of teacher who is tenured	-0.495†	-0.550†	-0.500†	-0.539†	-0.545†	-0.529†	-0.0728
	(0.298)	(0.303)	(0.297)	(0.302)	(0.301)	(0.290)	(0.0063)
ρ	-0.2883**	-0.0489	-0.2893**	-0.0691	-0.0737	-0.2832**	NA
	(0.0981)	(0.1226)	(0.0947)	(0.1093)	(0.0914)	(0.1024)	
σ	0.9644***	0.9764***	0.9724***	0.9626***	0.9687***	0.9697***	NA
	(0.0313)	(0.0231)	(0.0357)	(0.0300)	(0.0529)	(0.0319)	
F-Statistic on the excluded instrument	NA	NA	NA	NA	NA	NA	144.471
Observations	555	555	555	555	555	555	555

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. $\sigma^2 = Var(\varepsilon_i^j)$ and $\rho\sigma = Cov(\varepsilon_i^j, \eta_i^j)$.

Table 6: The Impacts of the RIECE Curriculum on Child Development

VARIABLES	General	GM	FM	RL	EL	PS	Number of Observations
One-level maximum sample							
RIECE curriculum (MLE)	0.540***‡	0.343*	0.261†	0.207	0.367*	0.362*	555
	(0.142)	(0.162)	(0.136)	(0.145)	(0.144)	(0.158)	
RIECE curriculum (2SLS)	0.473**‡	0.332*	0.307*	0.141	0.244†	0.302†	555
	(0.151)	(0.156)	(0.144)	(0.153)	(0.143)	(0.163)	
RIECE curriculum (OLS)	0.267**	0.298**‡	-0.0153	0.141	0.297**‡	0.0919	555
	(0.0957)	(0.0975)	(0.0961)	(0.100)	(0.0909)	(0.102)	
Two-level maximum sample							
RIECE curriculum (MLE)	0.653***‡	0.464*	0.334†	0.411**	0.520**	0.232	323
	(0.171)	(0.198)	(0.171)	(0.157)	(0.177)	(0.213)	
RIECE curriculum (2SLS)	0.508**	0.368†	0.404*	0.244	0.340†	0.139	323
	(0.188)	(0.193)	(0.186)	(0.177)	(0.181)	(0.210)	
RIECE curriculum (OLS)	0.342**	0.294*	-0.0182	0.327**	0.340**	0.139	323
	(0.122)	(0.130)	(0.130)	(0.121)	(0.121)	(0.127)	
Whole sample							
RIECE curriculum (MLE)	0.442***‡	0.354**	0.103	0.255*	0.323*	0.271†	718
	(0.120)	(0.136)	(0.121)	(0.121)	(0.139)	(0.139)	
RIECE curriculum (2SLS)	0.408**‡	0.382**	0.153	0.183	0.184	0.228	718
	(0.133)	(0.138)	(0.123)	(0.130)	(0.144)	(0.156)	
RIECE curriculum (OLS)	0.233**	0.288**‡	-0.0968	0.193*	0.269**	0.0963	718
	(0.0882)	(0.0873)	(0.0889)	(0.0865)	(0.0939)	(0.101)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$, include student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamins, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. The control variables for classroom, $\mathbf{Z}^j$, include student-teacher ratio including additional teacher, age level dummies, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. ‡ corresponds significance at the 5% level of multiple hypothesis correction for p-values using the Holm-Benferroni method with $\mathbf{X}_i^j$ as conditioning variables.

Table 7: The Impacts of the RIECE Curriculum on Child Development: Clustered by SAO, Clustered by Center, Drop Co-taught Classrooms, and the Randomization Outcome

VARIABLES	General	GM	FM	RL	EL	PS	Number of Observations
One-level maximum sample: clustered by SAO							
RIECE curriculum (MLE)	0.540**	0.343†	0.261	0.207	0.367*	0.362*	555
	(0.193)	(0.183)	(0.196)	(0.170)	(0.149)	(0.161)	
RIECE curriculum (2SLS)	0.473*	0.332*	0.307	0.141	0.244	0.302†	555
	(0.209)	(0.167)	(0.216)	(0.176)	(0.157)	(0.174)	
RIECE curriculum (OLS)	0.267†	0.298*	-0.0153	0.141	0.297*	0.0919	555
	(0.130)	(0.108)	(0.110)	(0.0981)	(0.114)	(0.117)	
One-level maximum sample: clustered by center							
RIECE curriculum (MLE)	0.540**	0.343†	0.261	0.207	0.367*	0.362*	555
	(0.192)	(0.184)	(0.196)	(0.171)	(0.155)	(0.168)	
RIECE curriculum (2SLS)	0.473*	0.332†	0.307	0.141	0.244	0.302†	555
	(0.210)	(0.172)	(0.209)	(0.176)	(0.157)	(0.175)	
RIECE curriculum (OLS)	0.267*	0.298*	-0.0153	0.141	0.297*	0.0919	555
	(0.118)	(0.117)	(0.115)	(0.112)	(0.114)	(0.101)	
One-level maximum sample: dropped co-taught classroom							
RIECE curriculum (MLE)	0.708***	0.396	0.532***	0.260	0.373	0.439†	429
	(0.200)	(0.258)	(0.156)	(0.221)	(0.246)	(0.244)	
RIECE curriculum (2SLS)	0.731**	0.359	0.693***	0.197	0.313	0.419†	429
	(0.223)	(0.220)	(0.209)	(0.219)	(0.209)	(0.239)	
RIECE curriculum (OLS)	0.232†	0.294*	0.107	0.0357	0.241*	0.00710	429
	(0.118)	(0.115)	(0.122)	(0.123)	(0.103)	(0.120)	
Randomization outcome (RT^j)							
Randomization outcome (OLS): One-level maximum sample	0.406***	0.284*	0.192†	0.149	0.260*	0.251*	555
	(0.113)	(0.115)	(0.106)	(0.117)	(0.114)	(0.118)	
Randomization outcome (OLS): Two-level maximum sample	0.523***	0.385*	0.296*	0.286†	0.365*	0.174	323
	(0.149)	(0.150)	(0.144)	(0.148)	(0.160)	(0.167)	
Randomization outcome (OLS): Whole sample	0.347***	0.327**	0.0826	0.167†	0.177	0.172	718
	(0.101)	(0.104)	(0.0913)	(0.0990)	(0.113)	(0.115)	

*** $p<0.001$, ** $p<0.01$, * $p<0.05$, † $p<0.1$. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$, include student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamins, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. The control variables for classroom, $\mathbf{Z}^j$, include student-teacher ratio including additional teacher, age level dummies, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

Table 8: The Impacts of Curriculum Exposure Period on Child Development

VARIABLES	General	GM	FM	RL	EL	PS	Number of Observations
One-level maximum sample							
Curriculum Exposure Period (2SLS)	0.0626**	0.0497*	0.0388†	0.0101	0.0345	0.0433†	551
	(0.0237)	(0.0246)	(0.0230)	(0.0236)	(0.0219)	(0.0254)	
Curriculum Exposure Period (OLS)	0.0291†	0.0279†	-0.00629	0.0231	0.0336*	0.0142	551
	(0.0151)	(0.0155)	(0.0150)	(0.0149)	(0.0146)	(0.0158)	
Two-level maximum sample							
Curriculum Exposure Period (2SLS)	0.0570†	0.0508	0.0503	0.0224	0.0446	0.00702	319
	(0.0303)	(0.0313)	(0.0315)	(0.0284)	(0.0276)	(0.0330)	
Curriculum Exposure Period (OLS)	0.0433*	0.0314	-0.0134	0.0488**	0.0521**	0.0255	319
	(0.0194)	(0.0214)	(0.0213)	(0.0187)	(0.0194)	(0.0209)	
Whole sample							
Curriculum Exposure Period (2SLS)	0.0557**	0.0558**	0.0187	0.0222	0.0267	0.0334	714
	(0.0206)	(0.0215)	(0.0193)	(0.0200)	(0.0216)	(0.0237)	
Curriculum Exposure Period (OLS)	0.0268†	0.0324*	-0.0192	0.0271*	0.0340*	0.0164	714
	(0.0137)	(0.0136)	(0.0144)	(0.0126)	(0.0147)	(0.0156)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$, include student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamins, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. The control variables for classroom, $\mathbf{Z}^j$, include student-teacher ratio including additional teacher, age level dummies, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

Table 9: Heterogeneous Effects of the RIECE Curriculum on Child Development

VARIABLES	General	GM	FM	RL	EL	PS	Number of Observations
(1) Children characteristics							
RIECE x boy dummy	0.0346	-0.0525	-0.105	-0.0129	-0.137	0.353*	555
	(0.170)	(0.173)	(0.177)	(0.172)	(0.167)	(0.163)	
RIECE x low birth weight dummy	0.118	-0.101	-0.106	0.244	0.266	0.0772	555
	(0.338)	(0.317)	(0.286)	(0.352)	(0.323)	(0.294)	
RIECE x chronic dummy	0.0930	0.338	0.246	-0.313	-0.0668	0.0535	555
	(0.283)	(0.243)	(0.259)	(0.314)	(0.277)	(0.235)	
(2) Household characteristics							
RIECE x household wealth index	-0.0741	0.00303	-0.138	-0.0420	-0.00285	-0.00546	507
	(0.104)	(0.100)	(0.122)	(0.0886)	(0.0908)	(0.100)	
RIECE x parental absence dummy	0.404*	0.00218	0.243	0.211	0.359*	0.340*	554
	(0.169)	(0.176)	(0.179)	(0.174)	(0.162)	(0.170)	
RIECE x parental marital status dummy	0.200	0.00318	0.137	0.247	0.270	-0.00999	553
	(0.184)	(0.184)	(0.186)	(0.183)	(0.172)	(0.188)	
(3) Main caregiver characteristics							
RIECE x MC's education	-0.0887	0.0702	0.0953	-0.0763	-0.0754	-0.246	498
	(0.192)	(0.205)	(0.214)	(0.190)	(0.162)	(0.198)	
RIECE x MC's age	0.00660	0.00872	0.00357	0.00508	0.000367	0.00199	495
	(0.00650)	(0.00711)	(0.00737)	(0.00617)	(0.00688)	(0.00669)	
RIECE x MC's digit-span recall score	-0.193**	-0.0947	-0.131†	-0.0919	-0.135†	-0.0968	355
	(0.0720)	(0.0712)	(0.0732)	(0.0760)	(0.0706)	(0.0691)	
(4) Mother characteristics							
RIECE x mother's education	0.0161	0.219	0.399†	-0.269	-0.0459	-0.169	261
	(0.216)	(0.254)	(0.220)	(0.235)	(0.245)	(0.234)	
RIECE x mother's age	0.0312†	0.0332†	0.0339†	0.0145	0.0131	0.000927	249
	(0.0161)	(0.0185)	(0.0184)	(0.0203)	(0.0177)	(0.0177)	
RIECE x Teenage pregnancy mother dummy	-0.306	-0.198	-0.607*	-0.110	-0.436†	0.362	249
	(0.261)	(0.327)	(0.290)	(0.328)	(0.254)	(0.300)	
(5) Teacher characteristics							
RIECE x average teacher's experience	0.0195	-0.0178	0.0352*	0.000878	0.0187	0.0163	555
	(0.0141)	(0.0144)	(0.0137)	(0.0143)	(0.0114)	(0.0147)	
RIECE x fraction of tenured teachers in the classroom	0.560*	0.00755	0.613*	0.383	0.386†	0.133	555
	(0.239)	(0.236)	(0.241)	(0.250)	(0.199)	(0.250)	
RIECE x fraction of teachers with bachelor degrees in early childhood education	0.369	0.103	0.466	0.354	-0.0495	0.173	555
	(0.249)	(0.266)	(0.294)	(0.268)	(0.222)	(0.281)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$ and $\mathbf{Z}^j$, are the same as the benchmark case in Table 6.

Table 10: The Impacts of Adoption Quality on Child Development

VARIABLES	General	GM	FM	RL	EL	PS	Number of Observations
One-level maximum sample							
Adoption quality (2SLS)	0.590**	0.413*	0.364*	0.186	0.312†	0.376†	555
	(0.182)	(0.185)	(0.172)	(0.185)	(0.173)	(0.196)	
Adoption quality (OLS)	0.367***	0.404***	0.0910	0.197†	0.307**	0.0837	555
	(0.111)	(0.115)	(0.112)	(0.117)	(0.105)	(0.117)	
Two-level maximum sample							
Adoption quality (2SLS)	0.697**	0.535*	0.493*	0.344	0.492*	0.186	323
	(0.243)	(0.247)	(0.238)	(0.230)	(0.238)	(0.271)	
Adoption quality (OLS)	0.424**	0.337*	0.112	0.418**	0.330*	0.0842	323
	(0.147)	(0.156)	(0.153)	(0.143)	(0.145)	(0.151)	
Whole sample							
Adoption quality (2SLS)	0.499**	0.467**	0.182	0.225	0.224	0.276	718
	(0.159)	(0.165)	(0.146)	(0.156)	(0.173)	(0.187)	
Adoption quality (OLS)	0.344**	0.416***	-0.0360	0.280**	0.279*	0.0775	718
	(0.105)	(0.104)	(0.106)	(0.102)	(0.110)	(0.119)	

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$, include student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamins, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. The control variables for classroom, $\mathbf{Z}^j$, include student-teacher ratio including additional teacher, age level dummies, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

Table 11: The Marginal Effect of the RIECE Curriculum on the Likelihood of Passing the Test (One-level Maximum Sample)

VARIABLES	General	GM	FM	RL	EL	PS
RIECE curriculum dummy	0.1130	0.0070	0.0341	0.0484	0.0417	0.1367*
	(0.0698)	(0.0648)	(0.0799)	(0.0512)	(0.0387)	(0.0639)
Child's age	0.2471***	-0.0093	0.3472***	0.0286	-0.0149	0.1054**
	(0.0610)	(0.0444)	(0.0658)	(0.0325)	(0.0243)	(0.0402)
Child's age squared	-0.0029***	-0.0001	-0.0041***	-0.0003	0.0002	-0.0012**
	(0.0007)	(0.0005)	(0.0007)	(0.0004)	(0.0003)	(0.0005)
Boy dummy	-0.0504	-0.0102	-0.0297	-0.0232	-0.0139	-0.1318***
	(0.0389)	(0.0341)	(0.0435)	(0.0280)	(0.0207)	(0.0338)
Low birth weight dummy	-0.0799	-0.0653	-0.1261†	-0.1327*	-0.0450	-0.0104
	(0.0666)	(0.0645)	(0.0738)	(0.0662)	(0.0471)	(0.0615)
Sibling dummy	0.0500	0.0071	0.0620	-0.0021	0.0172	0.0630†
	(0.0400)	(0.0341)	(0.0445)	(0.0283)	(0.0199)	(0.0343)
Chronic disease dummy	-0.0113	0.0366	-0.1006	-0.1341*	-0.0250	0.0202
	(0.0598)	(0.0492)	(0.0645)	(0.0587)	(0.0386)	(0.0509)
Student-teacher ratio	0.0024	0.0024	-0.0031	-0.0030	0.0076*	0.0007
	(0.0050)	(0.0043)	(0.0056)	(0.0035)	(0.0035)	(0.0043)
Teacher's experience	-0.0044	0.0025	0.0019	0.0002	-0.0016	0.0038
	(0.0049)	(0.0048)	(0.0056)	(0.0035)	(0.0031)	(0.0044)
Teacher's age	-0.0095	-0.0112	0.0040	0.0108	-0.0033	-0.0339
	(0.0248)	(0.0217)	(0.0280)	(0.0161)	(0.0152)	(0.0219)
Teacher's field	0.0256	0.0815	0.1904*	-0.0821	0.0095	-0.0089
	(0.0782)	(0.0637)	(0.0916)	(0.0521)	(0.0406)	(0.0634)
Number of Observations	555	555	555	555	555	555

*** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Robust standard errors are in parentheses. The control variables, $\mathbf{X}_i^j$, include student-teacher ratio including additional teacher, child's age, child's age squared, a dummy for being a boy, a dummy for having low birth weight, a dummy for having at least one sibling in the household, a dummy for having a chronic disease, a dummy for taking additional vitamins, a dummy for having lego at home, a dummy for having at least one jigsaw puzzle at home, a dummy for having at least one plastic/wooden shape sorter toy at home, a dummy for having at least one set of clay at home, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers. The control variables for classroom, $\mathbf{Z}^j$, include student-teacher ratio including additional teacher, age level dummies, the average teacher's age in each classroom, the average of teacher's age in each classroom squared, the average teacher's experience in each classroom, fraction of teachers with bachelor degrees in early childhood, and fraction of tenured teachers.

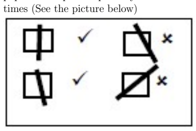
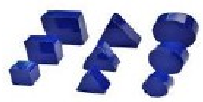
Table 12: The Number of DSPM Items over 19 Age Ranges

Age range	Gross Motor (GM)	Fine Motor (FM)	Receptive Language (RL)	Expressive Language (EL)	Personal and Social (PS)
Birth- 1 month	1 item	1 item	1 item	1 item	1 item
1-2 months	1 item	1 item	1 item	1 item	1 item
3-4 months	1 item	1 item	1 item	1 item	1 item
5-6 months	1 item	1 item	1 item	1 item	1 item
7-9 months	2 items	1 item	1 item	1 item	1 item
9 months	2 items	2 items	1 item	2 items	1 item
10-12 months	1 item	1 item	1 item	1 item	1 item
13-15 months	1 item	1 item	1 item	1 item	1 item
16-18 months	1 item	1 item	1 item	1 item	1 item
18 months	2 items	2 items	2 items	2 items	2 items
19-24 months	1 item	1 item	1 item	1 item	1 item
25-29 months	1 item	1 item	1 item	1 item	1 item
30 months	2 items	2 items	2 items	1 item	2 items
31-36 months	1 item	1 item	1 item	1 item	1 item
37-41 months	1 item	1 item	1 item	1 item	2 items
42 months	2 items	3 items	2 items	2 items	2 items
43-48 months	1 item	2 items	1 item	1 item	1 item
49-54 months	1 item	1 item	1 item	1 item	1 item
55-60 months	1 item	1 item	1 item	1 item	1 item

Table 13: Description of Key Variables for the Estimation of Heterogeneous Effects

Variable	Description
Low birth weight dummy	A dummy for low birth weight equals one if the child's birth weight was below 2,500 grams, and zero (low-educated) otherwise.
Chronic dummy	A dummy for chronic diseases equals one if the child has had asthma, allergies, thalassaemia, glucose-6-phosphate dehydrogenase deficiency (G6PD), anemia, heart disease, epilepsy, tonsillitis, lymphadenitis, pneumonopathy, enteropathy, mycosis, or nephropathy during the last 12 months of the interview, and zero (low-educated) otherwise.
Household wealth index	Household wealth index was constructed by running a Principal Component Analysis (PCA) using asset holdings: the number of houses/buildings, barns, huts, bicycles, motorcycles, cars, vans/pick-up trucks, motorized carts, farm tractors, four-wheel tractors, trucks, boats, boats with a small motor, telephones, computers/laptops, printers, tablets, air conditioners, cable TVs/satellite dishes, washing machines, televisions, refrigerators, and microwaves.
Parental absence dummy	A dummy for parental absence equals one if the child lives with neither parent present at home, and zero (low-educated) otherwise.
Parental marital status dummy	A dummy for parental marital status equals one if the parental status is divorced, separated, or widowed, and zero (low-educated) otherwise.
Main caregiver's education	A dummy for main caregiver's education that equals one if the main caregiver went beyond 9th grade or has more than nine years of schooling (highly-educated), and zero (low-educated) otherwise.
Mother's education	A dummy for mother's education that equals one if the mother went beyond 9th grade or has more than nine years of schooling (highly-educated), and zero (low-educated) otherwise.
Teenage pregnancy mother dummy	A dummy for teenage pregnancy mother is the mother who was pregnant at age between 10 to 19 years old, and zero (low-educated) otherwise.

Table 14: Examples of the Developmental Surveillance and Promotion Manual (DSPM)

Age	Gross Motor (GM)	Fine Motor (FM)	Receptive Language (RL)	Expressive Language (EL)	Personal and Social (PS)
43-48 months	101. Jumping on one leg continuously at least 2 times. (Tester may demonstrate) Pass: If the child can continuously move forward while jumping on one leg at least 2 times. ☐ Pass ☐ Fail	102. Cut 2 square pieces of paper with a size of 10 cm. (Tester may demonstrate) Pass: If the child uses the scissors to cut the paper into 2 parts separately at least 1 of 3 times (See the picture below)  Tools: 1. A pair of scissors. 2. A square piece of paper with a size of 10 cm. Note: Demonstrate every time, stop if the child cannot do. ☐ Pass ☐ Fail 103. Copy the positive sign (+) by drawing. (Tester must always demonstrate) Pass: If the child can copy the positive sign (the vertical line intersects the horizontal line) at least 1 of 3 times. Tools: 1. Pencil 2. Paper Note: The size of the child's drawing is not necessarily the same as the size of the sample. ☐ Pass ☐ Fail	104. Identify which object is bigger/smaller. The evaluation: Point to the medium-sized object. Then ask the child "Which object is bigger than this?" and "Which object is smaller than this?". Repeat for 3 sets of objects, starting and then asks the child that "which one is bigger than this" "which one is smaller than this". Ask all 3 sets of objects: circle, rectangle, and triangle. Pass: If the child can correctly answer 2 of 3 sets of objects. Tools: 3 sets of different objects of the same color.  ☐ Pass ☐ Fail	105. Speak at least 3 consecutive words in different contexts. (If cannot observe the child, ask either parent or teacher) The evaluation: Observe whether the child can communicate in 5 different contexts as follows: 1. Goodbye, e.g., "See you later." 2. Greeting, e.g., "Hello, mother." 3.Thankfulness, e.g., "Thank you teacher." 4.Opinion, e.g., "I think this one is more beautiful." 5.Apology, e.g., "I am sorry." Pass: Children can speak at least 3 consecutive words in all 5 contexts. Observe/ask from ☐ Children ☐ Teacher ☐ Parent ☐ Pass ☐ Fail	106. Put on 3 big buttons with a size of 2 cm by himself. (Tester may demonstrate) The evaluation: Put on and remove the buttons, and then let the child do it by himself. Pass: If the child can put on 3 buttons by himself. Tools: A rag doll that has three buttons of at least 2 cm. Note: No restriction of the sequence. ☐ Pass ☐ Fail
49-54 Months	107. Jumping on two legs to the left, the right, and backward continuously. (Tester may demonstrate) Pass: If the child can move to the left, the right, and the reverse continuously while jumping on two legs. ☐ Pass ☐ Fail	108. Assemble the parts of the pictures that were cut into 8 pieces. (The tester must show the child the completed picture before scrambling the 8 pieces.) Pass: If the child can assemble all 8 pieces correctly. Tools: 1 picture which is cut into 8 pieces.  ☐ Pass ☐ Fail	109. Select the pictures that represent day and night. (The tester must alternate the picture) Pass: If the child can point to the correct picture in 2 of 3 sets. Tools: Pictures 1. 3 pictures of day 2. 3 pictures of night ☐ Pass ☐ Fail	110. The child gives a reasonable response when he is asked "What will you do when you feel hot/sick/hungry/cold?" Pass: If the child can correctly answer 2 of 3 questions. (Fill the child's answers) Hot .... Sick .... Hungry .... Cold .... ☐ Pass ☐ Fail	111. Cleans himself after defecating (both the anus and hand). Pass: If the child can clean his anus and hand by himself after defecating. Observe/ask from ☐ Children ☐ Teacher ☐ Parent ☐ Pass ☐ Fail
55-60 months	112. Walking on heels. (Tester may demonstrate) Pass: If the child can walk forward on his heels for 4 steps without losing balance. Note: The child can extend his arms for balance while walking. ☐ Pass ☐ Fail	113. Holds a pencil correctly. (Tester does not demonstrate) The evaluation: Give a piece of paper and a pencil to the child, and then tell child to "write your name." Pass: If the child holds a pencil approximately 1-2 cm above the tip, and the pencil is grasped between the thumb, forefinger, and middle finger. Tools: 1. Pencil 2. Paper ☐ Pass ☐ Fail	114. Choosing 8 colors according to the order. Pass: If the child can pick up 8 color blocks correctly, according to the order. The order of color If correct fill √ in ☐ wrong fill X in ☐ 1. Red ☐ 2. Green ☐ 3. Blue ☐ 4. Black ☐ 5. White ☐ 6. Pink ☐ 7. Orange ☐ 8. Yellow ☐ Tools: 10 blocks with different colors Note: Do not order "Violet and Brown" ☐ Pass ☐ Fail	115. Take turns talking in a group. (If cannot observe, ask either parent or teacher) Pass: If the child can take turns talking in a group. Observe/ask from ☐ Children ☐ Teacher ☐ Parent ☐ Pass ☐ Fail	116. Play the role of an adult. (If cannot observe, ask either parent or teacher) Pass: The child can play the role of an adult, e.g., father, mother, teacher, doctor, nurse, group head, by mimicking the tone, action, attire. Observe/ask from ☐ Children ☐ Teacher ☐ Parent ☐ Pass ☐ Fail