

The Role of Preschool Quality in Promoting Child Development

Evidence from Rural Indonesia

Sally Anne Brinkman

Amer Hasan

Haeil Jung

Angela Kinnell

Nozomi Nakajima

Menno Pradhan



WORLD BANK GROUP

Education Global Practice Group

January 2016

Abstract

This paper reports on the quality of early childhood education in rural Indonesia. On average, the paper finds that centers created under the Indonesia Early Childhood Education and Development Project provide higher quality services than other types of preschools, as measured by a comprehensive instrument of preschool quality based on direct observation of classrooms in session (the Early Childhood Environment Rating Scale-Revised). The paper also examines the relationship between preschool quality and children's early development using three commonly applied measures of quality: (i) the Early Childhood Environment Rating Scale-Revised; (ii) teacher characteristics; and (iii) structural characteristics of preschool services, such as their size and amount of class time. First, correcting for measurement error using an instrumental variables approach, the findings suggest that preschool quality is a significant and meaningful positive predictor of children's developmental outcomes. Second, the findings for teacher characteristics

are mixed, suggesting that policies focused solely on hiring teachers based on experience and training will be insufficient to improve children's learning. Instead, policies must address the quality of professional development activities for teachers. Third, the amount of class time spent in early childhood programs is a significant positive predictor of children's developmental outcomes. This suggests that in rural Indonesia—where early childhood programs are relatively low dose—children are likely to benefit from attending longer hours of preschool, either playgroups or kindergartens. Lastly, the paper compares items in the Early Childhood Environment Rating Scale-Revised with Indonesia's national minimum service standards for early childhood education and development, and finds that the relationship between this alternative, context-appropriate measure of preschool quality and children's development outcomes strongly corroborates the earlier conclusions.

This paper is a product of the Education Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at ahasan1@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

The Role of Preschool Quality in Promoting Child Development: Evidence from Rural Indonesia

Sally Anne Brinkman¹, Amer Hasan², Haeil Jung³, Angela Kinnell⁴, Nozomi Nakajima², Menno Pradhan^{5*}

JEL classification codes: I20, I21

Keywords: Early Childhood Education and Development (ECED), quality, playgroups, kindergartens, classrooms, Indonesia

¹ Fraser Mustard Centre, Telethon Kids Institute, University of Western Australia, Perth, Australia.

² Education Global Practice, World Bank, Washington, DC, USA.

³ Department of Public Administration, Korea University, Seoul, Korea.

⁴ The University of Adelaide, Adelaide, Australia

⁵ University of Amsterdam and VU University, Amsterdam, the Netherlands.

* The authors are listed alphabetically to indicate that this paper is a team product in which all the authors contributed in different ways. We would like to thank Dedy Junaedi, Upik Sabainingrum, Anas Sutisna, Lulus Kusbudiharjo and Mulyana for managing the fieldwork. Mayla Safuro Lestari Putri and Mulyana provided excellent research assistance. Data collection was partially funded by the Government of the Kingdom of the Netherlands through the Dutch Education Support Program Trust Fund (TF057272).

1. Introduction

The quality of preschool education⁶ plays an important role in early development. Research to date has shown that children who have high-quality early childhood education experiences have better cognitive and socio-emotional outcomes (Engle et al., 2011). Unfortunately, providing quality early childhood education services can be particularly challenging in poor, rural contexts – especially so in developing countries. As a result, few studies have been able to use an internationally comparable measure of quality to examine the link between quality and child development outcomes in remote, resource-constrained environments.

This study investigates the quality of early childhood education programs in 303 villages in Indonesia. We use data collected in 2013 as part of an impact evaluation of the Indonesia Early Childhood Education and Development (ECED) Project, which provides rich information on early childhood classrooms, teacher characteristics, and child development outcomes.

Indonesia offers a useful setting to explore early childhood education as the government has been highly committed to improving both access to and quality of preschools in the country (Hasan, Hyson and Chang, 2013). In 2009, the Ministry of Education and Culture established national standards for ECED quality, further emphasizing the importance of investing in high-quality programs. By providing an overview of the quality of early childhood education services in poor, rural areas in Indonesia, this study will help inform policy makers whether the current national quality standards align well with desired child development outcomes.

This paper is organized as follows. Section 2 reviews the literature on early childhood education quality by discussing three components of quality: observed preschool quality (focusing on the ECERS-R, an internationally comparable measure), teacher characteristics (measured using their education and experience), and structural characteristics. Section 3 provides an overview of the early childhood education policy landscape in Indonesia and the research questions in this study. Then, section 4 introduces the data and methods of the study. In section 5, we outline the empirical strategy – including approaches to correcting for measurement error. Results are presented in section 6. Section 7 introduces an alternative measure of classroom quality that is more context-specific to Indonesia and presents the association between this alternative measure of preschool quality and children’s development. It also puts the quality of Indonesian ECED services in context – both internationally and in relation to local standards. We discuss our findings and conclude in section 8.

2. Literature review

The following review of the literature discusses empirical research on the link between preschool quality and children’s early childhood developmental outcomes. We focus on three aspects of preschool quality: observed preschool quality, teacher characteristics, and other structural characteristics of early childhood programs.

Observed preschool quality

Decades of research on child development have established the benefits of providing high-quality early childhood programs to young children (Campbell and Ramey, 1994; NICHD, 2005). One of the most widely used observational measures of quality of early childhood classroom environments is the revised version of the Early Childhood Environment Rating Scale (ECERS-

⁶ In this paper, the terms “preschool” and “early childhood education” are used interchangeably, and refer broadly to organized, center-based early learning environments for children from the ages of 3 until they enter primary school – usually playgroups and kindergartens.

R) (Harms, Clifford and Cryer, 2005).

Empirical studies using ECERS-R have generally found positive associations between the quality of early childhood programs and child development. These measures of child development range from language and cognitive skills (Burchinal et al., 2008; Peisner-Feinberg and Burchinal, 1997; Peisner-Feinberg et al., 2001) to social-behavioral development (Burchinal et al., 2008; Sylva et al., 2006). However, the magnitude of the correlations between observed classroom quality and children's developmental outcomes has been small—ranging from 0.04 standard deviations (Duncan, 2003) to 0.18 standard deviations (Peisner-Feinberg et al., 2001). These small associations, however, are not specific to the ECERS-R alone⁷ (Gordon et al., 2013; Vandell and Wolfe, 2000). Other widely-used measures of observed classroom quality – such as the Classroom Assessment Scoring System (CLASS) – have also shown fairly small associations with child development outcomes.⁸ In studies that have produced particularly small effect sizes, authors have discussed the idea that effect sizes of observed classroom quality are likely to be underestimated due to measurement error (Duncan, 2003). This suggests the need for further work on addressing the measurement error present in widely-used observational measures of classroom quality like the ECERS-R.

Prior studies on the association between preschool quality and child development also raise concerns of external validity because they have often focused on high-quality, “model” early childhood education programs in the U.S. (Barnett, 1995; Magnuson et al., 2007; Duncan and Magnuson, 2013). Unfortunately, most children in the world do not have access to such high-quality programs. As a result, it is important to know whether the positive relationship between preschool quality and child development holds true in other contexts. Studies conducted in the U.S. have looked at the impact of a range of preschool classroom quality on child development (Magnuson et al., 2007; Phillips et al., 2001) but it is still rare to find evidence from poor, rural areas in developing countries. Recent studies conducted in Bangladesh (Aboud, 2006), India (Rao, 2010), Kenya, Zanzibar and Uganda (Malmberg et al., 2011) found positive correlations between ECERS-R scores and children's development outcomes. However, with the exception of an East African study (Malmberg et al., 2011), analyses on the link between observational measures of the quality of the educational environment and child development outcomes have not introduced extensive child, family, and community-level covariates to control for confounding factors. Given the increasing focus on early childhood education interventions in developing countries, there is a need for more rigorous empirical evidence on this topic.

Teacher characteristics

In addition to directly observing the quality of the preschool learning environment, teacher characteristics are another common way of measuring the quality of early childhood programs. There is considerable policy interest in understanding the link between teacher qualifications and children's developmental outcomes, since governments can more readily regulate a teacher's level of education, experience, and training than observed classroom quality. In addition, teacher qualifications are often linked to teacher salary, which tends to be the largest expenditure in early childhood education programs in developing countries (Early et al., 2007; Glewwe and Kremer,

⁷ For example, the effect size for the Classroom Assessment Scoring System (CLASS) typically ranges from 0.06 to 0.10 standard deviations (Howes et al., 2008). For an overview of the associations between various measure of classroom quality and child development, see Table 2 in Vandell and Wolfe (2000).

⁸ The Classroom Assessment Scoring System (CLASS) is an observational measure of classroom quality similar in many ways to the ECERS-R.

2006). As a result, early childhood education policy makers often face the challenge of implementing a balanced teacher qualification standard that helps recruit teachers who can produce high-quality child development outcomes but also ensures that teacher salaries are affordable.

However, prior research on the link between teacher characteristics in early childhood programs and children's developmental outcomes is not uniformly positive. Some studies show that classrooms with more highly educated and trained teachers are associated with higher quality care (Burchinal et al., 2002; Phillips et al., 2001). In contrast, other studies have demonstrated that once unobserved differences across centers are controlled for using center fixed effects, the effect of teacher's education and training on child development outcomes disappear (Blau, 2000; Early et al., 2006, Early et al., 2007). These mixed findings on teacher characteristics are not unique to preschools; similar research findings are obtained from non-experimental research on primary and secondary school teachers (Early et al., 2007; Hanushek and Rivkin, 2006). Evidence from developing countries shows slightly stronger support for a positive relationship between teacher education and student achievement (Hanushek, 2003; Hanushek and Rivkin, 2006). However, one criticism of many of these studies is that they do not adequately control for omitted variables. Further analyses using extensive control variables—especially from low- and middle-income country contexts—can help clarify the inconclusive evidence that exists on this topic.

Structural characteristics

There are other characteristics of preschools that can perhaps be regulated even more easily than teacher characteristics. These so-called structural characteristics are usually easier to mandate and less costly to assess than observing the quality of individual classrooms. As a result, there is considerable policy interest in understanding whether structural characteristics (beyond teacher qualifications) matter in explaining children's early development.

Two structural characteristics that are often investigated in studies of preschool quality are teacher-child ratio and hours of operation. For teacher-child ratio, studies have generally shown that in lower-ratio settings, teachers spend less time managing children in the classroom and are able to provide more stimulating, responsive, and warm care (Phillips et al. 2001, Burchinal et al., 2000). However, Blau (2000), Mashburn et al. (2008) and Duncan and Magnuson (2013) find that the significant association between teacher-child ratio and child development outcomes disappears when regression models control for other characteristics of the preschool classroom (i.e., adding more classroom-level variables or using classroom fixed effects). This suggests that while teacher-child ratios matter, they may only affect children's development indirectly, by influencing their experiences with their teachers and peers within the classroom.

Another structural characteristic that has received attention in the literature is dose or duration of preschool, which is often measured by hours of operation. To date, findings on the relationship between hours spent in early childhood education and children's developmental outcomes are mixed. For cognitive and language development, studies find positive correlation between time spent in care and children's development but the association disappears once children's family background and other preschool aspects are controlled (Duncan, 2003; NICHD 2000; Vandell, 2004). Similarly, studies that have looked at children's social competence and behavioral problems have also found mixed results. In the US, researchers found that even when holding family factors constant, children who spend more hours in preschool have more behavioral problems than their peers who spend fewer hours in preschool settings (NICHD, 2003; Vandell, 2004). In contrast, more recent research from Norway finds little evidence of behavioral issues in children from spending more hours in child care (Zachrisson et al, 2013). To our knowledge, early

childhood education studies from developing settings have yet to examine whether structural characteristics matter to children's developmental outcomes even after controlling for key confounding factors such as observed classroom quality (which captures teacher and student interactions within classrooms) and teacher characteristics.

Thus, this study addresses several limitations in the prior literature on the relationship between the quality of early childhood programs and children's outcomes. First, we control for local characteristics by using district fixed effects. Second, we correct for possible measurement error in observational measures of classroom quality using an instrumental variable approach. Third, we contribute to the growing evidence base from developing countries on the predictive validity of observed classroom quality on children's developmental outcomes by introducing an extensive set of control variables to our models. Fourth, we contribute to the literature on the link between structural aspects of preschool quality (measured using teacher characteristics, teacher-child ratio, and hours of operation) and early developmental outcomes, which remains understudied in low- and middle-income country contexts.

3. Indonesian context

Several types of early childhood services exist in Indonesia; they serve different purposes and are administered by different ministries. Table 1 summarizes the key features of some of these programs.

Among the many existing early childhood programs in Indonesia, kindergartens, Islamic kindergartens, and playgroups focus on providing education directly to children. Generally, playgroups emphasize learning through play while kindergartens and Islamic kindergartens help prepare students for primary school (Hasan, Hyson and Chang, 2013). Given the focus on early childhood education, these three types of ECED programs will be the focus of our study.

In recent years, the government of Indonesia has prioritized early childhood education and development. The national standard for ECED (hereafter referred to as the Indonesia ECED Standard) was developed by the Ministry of Education and Culture in 2009 with the aim to set minimum standards around (i) children's development milestones by age; (ii) qualifications for teachers and education personnel; (iii) content, learning processes, and assessment; and (iv) facility, infrastructure, management, and financing.⁹ By establishing the national standard, the government has emphasized the need to ensure overall quality in early childhood programs across the country.

One of the key areas of the national standard is teacher qualifications. Teachers must have a university degree and assistant teachers must have at least a high school diploma, in addition to training in early childhood education from an accredited training provider. Teachers must also demonstrate a wide range of professional competencies in teaching young children, which would require significant experience in the classroom. As a result, the national standard outlines clear guidelines on the requirements for teacher's education, experience, and training in preschools.

In addition to teacher qualifications, the Indonesia ECED Standard defines policies on structural aspects of early childhood centers, such as class size and duration. In kindergartens, the maximum ratio of student to staff is 20 to 1. In playgroups and daycare centers, the maximum ratio depends on the age of the children in the class; for children ages 5-6, the ratio is 15 to 1 while for children less than 1 year-old, the ratio is 4 to 1. Similarly, the duration of programs depends on the age group. According to these standards, infants (age 0-2) are meant to attend for 2 hours a week,

⁹ As of late 2014 these standards have been updated. However, at the time data for this study was being collected, the 2009 version of the Indonesia standards were in effect. These are the standards described in this section.

toddlers (age 2-4) for 6 hours a week, children in kindergartens (age 4-6) are to attend 3 hours a day for 5 to 6 days per week, and those in playgroups or daycare (age 4-6) are meant to attend 3 hours a day for 3 days per week. Although these teacher and structural policies have been in place since 2009, little is known about whether and how strictly they are adhered to and whether they actually help promote child development outcomes.

As the government of Indonesia moves to expand and improve early childhood education programs across the country, there is immense interest among policy makers for evidence on whether classroom quality, teacher qualifications, and other structural characteristics (i.e., class size and duration)—all of which are outlined in the national ECED standard—do in fact support positive developmental outcomes for children. Thus, this study focuses on the following three research questions in the context of 303 poor villages in Indonesia:

- i. Does preschool quality—as measured by direct observation of preschool centers in session—predict children’s development outcomes?
- ii. Do teacher characteristics (averaged at the center level) predict children’s development?
and
- iii. Do structural characteristics of preschool centers—such as student-to-staff ratios and hours of operation—predict children’s developmental outcomes?

4. Data and Measures

Data

Data for this analysis was collected in 2013 as part of an impact evaluation of the Indonesia ECED Project (see Pradhan et al., 2013 for detailed study protocol). The Indonesia ECED Project aimed to improve poor children’s access to ECED services and enhance their school readiness. Through the project, participating villages created playgroups (hereafter referred to as *project playgroups*) and received training for teachers in these playgroups. As part of the project evaluation, data on children enrolled in project playgroups as well as their peers living in the same village attending other types of early childhood programs (i.e., kindergartens, Islamic kindergartens, and non-project playgroups) were collected. The quality of 578 early childhood programs, located in 303 poor villages across 9 districts, was observed during the 2013 survey conducted as part of the impact evaluation of the Indonesia ECED Project.¹⁰ In each of these centers child development outcomes of approximately 15 students were measured.¹¹

Measures

Early Development Instrument (EDI)

The key dependent variables of interest are children’s developmental outcomes measured using the Early Development Instrument (EDI), which has been shown to be a valid and reliable measure of child development internationally (Ip et al., 2013; Janus et al., 2011; Janus et al., 2007; Janus and Offord, 2007; Brinkman et al., 2013). The EDI is comprised of five domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge. Each domain is scored from 1 (low) to 10 (high). A teacher in the early childhood education center in which the child was enrolled

¹⁰ There were 310 total villages sampled in the evaluation of the Indonesia ECED Project. In 7 of these villages, the available centers did not meet the study’s criteria used for sample inclusion. In the remaining 303 villages, two centers were typically observed. The sampling rule was to pick one project playgroup (when available) and one other center – either a kindergarten or a non-project playgroup.

¹¹ In practice, the number of students sampled in each center ranged from 4 to 16.

completed the child's EDI (though this teacher was not always the child's direct classroom teacher). The EDI was adapted and translated for use in the Indonesia ECED Project by the authors and members of the research team.¹² Descriptive statistics of the EDI for the analytic sample are shown in Table 2.

Overall, we find variation in the EDI outcomes of children in the sample. Children score relatively high on the physical health and wellbeing domain (mean score of 8.1 out of 10) while results for social competence, emotional maturity, language and cognitive development, and communication and general knowledge range between 6.0 and 6.8. We also find variation in the developmental outcomes across different types of centers. On average, the means score of all EDI domains are higher among children enrolled in kindergarten and Islamic kindergarten than those attending non-project playgroup and project-playgroup. This likely reflects the fact that children are older when they are in kindergartens.

Child covariates

Along with the EDI, the data contains information on basic demographic information about the children in the sample. Summary statistics in table 2 show that while both girls and boys are equally represented across the four types of early childhood programs, children attending kindergarten are slightly older (mean age of 5 years) than those enrolled in playgroups (mean age of 4 years). Islamic Kindergartens have slightly higher rates of special needs children than both kindergartens and playgroups. The highest level of education attained by mothers of children in the sample is similar across all four types of ECD; about 40% report they have completed elementary school, followed by junior high (about 25%), senior high (about 25%), and higher education (less than 10%).

ECERS-R

One of the key predictor variables of interest is the quality of early childhood programs as measured by ECERS-R—an observation-based assessment of the learning environment of early childhood centers. Seven subscales make up the ECERS-R: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. The instrument has good test-retest reliability, high inter-rater reliability (Clifford et al., 2010), and many studies have demonstrated its predictive validity (Burchinal et al., 2008; Montes et al., 2005; Peisner-Feinberg et al., 2001).

In accordance with the instructions of use for the ECERS-R, two raters assessed each center at the same time. Both raters were present in the room with the class they were observing for three hours and followed this group if they left the room for outdoor play. Raters did not interact with staff or students during their observation. Prior to observation in the field, all raters were provided with training on the use of the ECERS-R. This training involved viewing of the ECERS-R video training package, practice observations in settings similar to those that would be rated as part of the study, and subsequent review of these practice ratings.

Each item on the seven subscales was assessed by the two raters on a seven-point Likert

¹² During piloting of the initial adaptation of the EDI it was found that respondents struggled with the original 5 point Likert scale response options. As a result, response options were changed to the binary can (“bisa” in Indonesian language) or cannot (“tidak bisa”). It was also found that respondents found it difficult to respond to questions, which asked for their subjective evaluation of a child's abilities (e.g. How would you rate [child's name]'s overall social/emotional development?). Questions such as these were omitted from the final Indonesian version of the instrument. Finally, the wording of 11 questions was altered to make their meaning clearer in the Indonesian language.

scale, which ranged from inadequate (1), minimal (3), good (5) to excellent (7). Raters began by assessing whether the requirements for a score of 1 were met; if they were, the rater assessed whether the indicators for a score of 3 were met, and so on through the scale. The ‘stop-scoring’ method was used. Stop-scoring means that when all indicators for a certain score were not met, scoring was stopped and that score was awarded regardless of whether later indicators for higher scores were met. For each subscale, rater one and rater two’s scores are averaged to construct the score for each subscale. Then, scores on the seven subscales are averaged to provide a total (mean) ECERS-R score. Descriptive statistics of the ECERS-R are shown in Table 3.¹³

The overall mean score on the ECERS-R is 2.983 (out of 7). We find variation in scores across the seven subscales of the ECERS-R with the highest scores in interactions (4.069 units) and the lowest scores in activities (2.480 units). We also find a range of scores across different types of centers. On average, we find the highest total score in ECERS-R for project playgroups (3.089 units), followed by kindergartens (3.035 units), non-project playgroups (2.731 units), and Islamic kindergartens (2.609 units). These averages over the seven subscales mask substantial variation in quality between different types of services. For instance playgroups established under the project have higher observed quality than kindergartens on several dimensions such as space and furnishings, activities, and interactions. In contrast, kindergartens seem to do better than project playgroups in terms of the parents and staff and language-reasoning subscales.

Aggregate teacher characteristics

Variables on teacher qualification are aggregate characteristics at the center level. They are constructed using the information of teachers who were present on the day of the ECERS-R assessment. These teachers completed the student’s EDI but may not necessarily be the child’s classroom teacher. Each teacher variable is created by dividing the number of teachers in a center with a particular characteristic by the total number of teachers observed in that center on that day. In our sample of 578 centers, between 1 and 7 teachers were observed in each center. Thus, teacher characteristics in our analysis are continuous variables that measure the average teacher characteristics at the center-level.

Summary statistics of these teacher characteristics are described in Table 3. The majority of teachers in kindergartens (64.8%) and Islamic kindergartens (48.1%) have a post-secondary education degree, while only 27.3% in non-project playgroups and 23.3% in project playgroups have any post-secondary qualifications. Similarly, teachers in kindergartens and Islamic kindergartens have over 8 years of teaching experience, while teachers in playgroups only have an average of 5 to 6 years. Teachers were asked to report on past experience. Their responses fell in the following categories – they had been teachers in kindergartens or playgroups before, they had served as cadres for the integrated village health service (Posyandu), family planning or a woman’s group or they had no prior experience. About a third of teachers (29.9%) in the sample did not have any prior ECED experience. One of the key features of the Indonesia ECED Project was provision of training to ECD teachers. The training under the project (hereinafter referred to as *project training*) emphasized interactive learning, demonstrations, and role-play to become effective instructors for young children.¹⁴ The project training was 200 hours, which was considerably longer than training typically offered by accredited teacher training programs in Indonesia even today (Hasan, Hyson and Chang, 2013). As shown, 52.2% of teachers in project

¹³ It is important to note that while the range of the ECERS-R only goes from 1 to 7 – the differences in quality between a 3 and a 5 are substantial.

¹⁴ The extent to which this model was consistently followed is unclear.

playgroups received 200 hours of teacher training through the project while 10.1% of teachers received only 100 hours of project training. On average, 15.7% of teachers had never received any training in early childhood education.

Structural characteristics

Two center-level variables (student-to-staff ratio and hours of operation per week) are included to account for structural characteristics. Summary statistics of these structural characteristics are described in Table 3. Kindergartens have the highest student to teacher ratio, followed by project playgroups, Islamic kindergartens and non-project playgroup. In terms of dose or intensity of early childhood education, children enrolled in kindergartens and Islamic kindergartens spend on average 15 hours/week in ECD centers whereas their peers enrolled in playgroups spend on average 10-11 hours/week.

The total sample consists of 578 early childhood education centers with observations on child development outcomes for over 8,300 children.¹⁵ We have missing data on teacher and structural characteristics of the centers, which reduces our analytic sample size to 566 centers. This results in usable data on child development outcomes for 7,946 children.¹⁶

5. Empirical strategy

First, we examine the simple correlation between observed classroom quality using the ECERS-R and children’s developmental outcomes (Model 1). Next, we seek to understand whether ECERS-R predicts children’s developmental outcomes while controlling for child, center, and district characteristics (Model 2):

$$Y_{ijk} = \beta_0 + \beta_1 Q_{jk} + \beta_2 X_{ijk} + \beta_3 S_{jk} + \nu_k + u_{ij} \quad (1)$$

where Y_{ijk} is the developmental outcome (one of the EDI domains) for child i enrolled in center j located in district k , Q_{jk} is the mean ECERS-R score of center j located in district k , X_{ijk} represents a vector of observable child characteristics for child i in center j located in district k , and S_{jk} is a dummy variable indicating the type of early childhood service (1=kindergarten/Islamic kindergarten, 0=project/non-project playgroup) provided in center j located in district k . District-specific unobserved variables (using district fixed effects) are embedded in ν_k and the error term is u_{ij} . We are interested in the parameter β_1 , which measures the relationship between ECERS-R and child development outcomes while holding all other variables constant.

In Model 3, we add average teacher characteristics (T_j) at the center level to equation 1 in order to examine whether teacher characteristics predict child development outcomes, over and above observed classroom quality:

$$Y_{ijk} = \beta_0 + \beta_1 Q_{jk} + \beta_2 T_{jk} + \beta_3 X_{ijk} + \beta_4 S_{jk} + \nu_k + u_{ij} \quad (2)$$

We are interested in the parameter β_2 , which measures the relationship between teacher characteristics (i.e., level of education, years of teaching, ECED experience, and training) and child development outcomes, holding all other variables constant. We are also interested in whether β_1 in Model 3 changes—if at all—compared to β_1 in Model 2.

In Model 4, we add center-level variables (C_j) in order to examine whether other structural

¹⁵ The sample size for each EDI domain is slightly different. N=8,348 for physical health, 8,317 for social competence, 8,306 for emotional maturity, 8,319 for language and cognitive development, and 8,345 for communication and general knowledge.

¹⁶ Sample size for each EDI domain also varies for these 566 centers – going up to N=7,984 for emotional maturity.

characteristics (i.e., student-to-staff ratio and hours of operation per week) predict child development outcomes, while controlling for observed classroom quality and teacher characteristics.

$$Y_{ijk} = \beta_0 + \beta_1 Q_{jk} + \beta_2 T_{jk} + \beta_3 C_{jk} + \beta_4 X_{ijk} + \beta_4 S_{jk} + v_k + u_{ij} \quad (3)$$

We are interested in the parameter β_3 , which measures the relationship between structural characteristics and child development outcomes, holding all other variables constant. We are also interested in whether β_1 and β_2 in Model 4 change—if at all—compared to β_1 and β_2 in Model 3.

In Model 5, we use an instrumental variable approach to correct for possible measurement errors in our score of observed classroom quality. Such quality ratings are subject to measurement error as they involve judgment on the part of each observer. Measurement error increases the noise in the quality variable leading to a downwards bias in the estimated correlations between the quality and child development outcomes. To correct for this potential bias, we exploit the fact that two observers rated each center independently. This allows us to use an instrumental variable approach, where one observer's rating serves as an instrument for the rating of the other observer for the same center. If differences between the two raters are caused by measurement error, this method will eliminate the downward bias in the estimated correlation. Thus instead of using this average ECERS-R score as we have done in equations 1 to 3, we use the first rater's score Q_j^1 , as an instrument for the second rater's score Q_j^2 , in a 2 Stage Least Squares (2SLS) model as follows:

First stage regression:

$$Q_{jk}^2 = \alpha_0 + \alpha_1 Q_{jk}^1 + \alpha_2 T_{jk} + \alpha_3 C_{jk} + \alpha_4 X_{ijk} + \alpha_4 S_{jk} + v_k + u_{ij} \quad (4.1)$$

Second stage regression:

$$Y_{ijk} = \gamma_0 + \gamma_1 \widehat{Q}_{jk}^2 + \gamma_2 T_{jk} + \gamma_3 C_{jk} + \gamma_4 X_{ijk} + \gamma_4 S_{jk} + v_k + \varepsilon_{ij} \quad (4.2)$$

We are primarily interested in the parameter γ_1 , which measures the relationship between ECERS-R and child development outcomes.

6. Results

Tables 4 through 8 present the unstandardized (raw EDI scores) regression results of the five models. Each table shows the outputs for a different EDI domain and each column presents the results from a separate model. Below, we organize our results around our three measures of preschool quality—ECERS-R, aggregate teacher characteristics, and structural characteristics—and their association to child development outcomes.

Observed classroom quality using ECERS-R

The simple correlations (Model 1) of ECERS-R and child development outcomes show a significant positive relationship in four out of five EDI domains. A one-unit increase in the ECERS-R (which is on a scale of 1 to 7) is associated with a 0.107 unit increase in the physical health and well-being EDI (which is on a scale of 1 to 10). Similarly, the coefficient on ECERS-R is 0.150 for social competence, 0.141 for emotional maturity, and 0.129 for language and cognitive development (though it is statistically insignificant), and 0.151 for communication and general knowledge.

Results of Model 2 show that ECERS-R is a significant predictor of only one EDI domain, when controlling for a set of child characteristics and center type with district fixed effects. A one-unit increase in the ECERS-R is associated with a 0.096 unit increase in the social

competence domain, holding all else constant. The point estimates for the remaining domains are substantially smaller than in the case of Model 1 and statistically insignificant.

Once we add teacher characteristics in Model 3, we find that ECERS-R is no longer a significant predictor of the social competence domain. Given that the percent of teachers with senior secondary education and the percent of teachers with post-secondary education are significant predictors of social competence in Model 3, the decrease in the ECERS-R coefficient is likely due to collinearity between classroom quality and teacher characteristics. In Model 4, the addition of structural characteristics results in ECERS-R to be a significant predictor of physical health and well-being. The positive coefficient of the student-to-staff ratio variable in Model 4 suggests that it further absorbs some of the residual variability in the model.

Thus far our results in Models 1 through 4 have largely replicated the results found by previous studies. First, we find a positive correlation between observed classroom quality and children's developmental outcomes in the basic model with no covariates (Model 1), which is consistent with prior large-scale studies (Duncan, 2003; Peisner-Feinberg et al., 2001). However, the magnitude of the association is on the smaller end of the range of effect sizes found in these previous studies. The standardized beta regression coefficient¹⁷ on observed classroom quality is 0.072 to 0.100 standard deviations in our study, while the effect size of observed classroom quality in prior studies range from 0.11 standard deviations (Duncan, 2003) to 0.26 standard deviations (Peisner-Feinberg et al., 2001). Second, the null results in Models 2 to 4 are similar to previous studies that have found non-significant associations between observed center quality and child development when models introduce extensive controls for child, center, and environment characteristics (Gordon et al., 2013; Sylva et al., 2006). One of the limitations of observational measures of classroom quality is that its effect size is likely to be underestimated due to measurement error. As a result, we attempt to correct for measurement error to yield more precise estimates of observed classroom quality on children's developmental outcomes in the following model.

The results in Model 5 suggest that using an instrumental variables approach results in stronger associations between ECERS-R and children's developmental outcomes. Compared to Model 4, Model 5 shows that a one-unit increase in ECERS-R is associated with a 0.107 unit increase in the physical health and well-being domain, holding all else constant. Similarly, the coefficient of ECERS-R is 0.111 for the social competence domain while controlling for all other variables in the model. For other domains this same pattern is observed though the correlations are not statistically significant at the conventional levels.

Overall, the magnitude of the coefficient on ECERS-R changes in a similar pattern from Model 1 to 5 across the five EDI domains. The simple correlation of ECERS-R and EDI is between 0.107 and 0.151 (Model 1). However, introducing control variables into the regression in Models 2, 3, and 4 generally decreases the coefficient on ECERS-R. Lastly, moving from Model 4 to 5, we find the magnitude of the coefficient on ECERS-R increases when an instrumental variables approach is used to correct for measurement error in the ECERS-R instrument.

Teacher characteristics

For teacher characteristics aggregated at the center level, we generally find mixed results across the various models. In terms of teacher's education level, results from Models 3 through 5 show that increasing the percent of teachers with a secondary or a post-secondary degree predicts higher child developmental outcomes in social competence EDI and communication and general

¹⁷ Regression results for Models 1-5 with standardized beta coefficients are available upon request.

knowledge EDI, while controlling for all other factors such as observed classroom quality and structural characteristics. For example, a 100% increase in teachers with a post-secondary education degree is associated with a 1.073 to 1.103 unit increase in children's social competence, holding all else constant. For both social competence EDI and communication and general knowledge EDI, the coefficient on percent of teachers with post-secondary education is larger than that of senior secondary education. For the other three EDI domains, we find no significant relationship between increasing the share of teachers with higher levels of education and children's developmental outcomes.

Across the models, we find that increasing the mean years of teaching experience at the center level does not predict higher EDI scores. In fact, for the physical health and well-being domain and the communication and general knowledge domain, an additional year of average teaching experience in a center is associated with a 0.013 to 0.015 unit decrease (for physical health and well-being) or a 0.044 to 0.045 unit decrease (for communication and general knowledge), holding all else constant. In interpreting these results, it is important to keep in mind that these data are aggregated at the center level and do not link specific children to specific teachers.

For previous experience in ECED, we find null results on children's development. Similarly, we find no significant associations between teachers who have received teacher training and child development outcomes.

Structural characteristics

Over and above aspects of the quality of the learning environment, we find that the quantity of exposure (or dose) of early childhood education is important to children's developmental outcomes in two out of five EDI domains. An additional hour per week of early childhood education is associated with a 0.039 unit increase in a child's language and cognitive development and a 0.028 unit increase in social competence development, controlling for all other variables. In contrast, we find null results for student-to-staff ratio on children's developmental outcomes, except in physical health and well-being.

7. Reassessing the quality-outcome relationship using the Indonesia ECED standard

The results so far show that ECERS-R is a significant predictor of child development in two domains: physical health and well-being and social competence. While the ECERS-R is a useful measure to compare the quality of early childhood programs across various settings, it can be a challenge for policy makers to interpret the ECERS-R within a particular context. For example, a closer look at the items in the ECERS-R reveals that some items may be less relevant than others in our study's setting—poor, rural villages in Indonesia. Item 27 outlines the provision of TV, video or computers for classroom activities. Item 3 describes the provision of soft furnishings such as carpeted space and cushions for children's relaxation and comfort. While all of these provisions would be great to have, they are often unfeasible in rural, resource-constrained environments. In over 3 weeks of field observation, we rarely saw soft furnishings—sitting on the floor for children and adults alike is the norm in rural Indonesia – even when soft furnishings are available. We did encounter one center with a computer lab during piloting of the ECERS-R but this was a state-of-the-art center where not one but three international donors had channeled funding.

Indeed, recent studies recommend researchers and practitioners alike to pay more attention to how measures of quality like ECERS-R align with local regulations and accreditation standards (Gordon et al., 2013). In an effort to align the ECERS-R data with the reality of the Indonesian

context, we turn to the Indonesian national standard for ECED as an alternative way to look at quality. Using textual analysis, we compared ECERS-R to Indonesia's ECED Standard and found 28 out of 43 ECERS-R items discussed in the national standard (see Table 9).¹⁸

Based on this crosswalk of ECERS-R items to Indonesia's own standards, we conducted further analyses. First, we calculated the mean ECERS-R score using only the 28 items that were found to be common between ECERS-R and the national standard—we call this alternative measure of quality, the “Indonesia standard.” The mean and standard deviation of this alternative measure is presented in Table 10.

As shown in Figure 1, many of the items that were retained in the Indonesia standard are those that pick up variation in quality, such as items 16, 17, and 22. Meanwhile, the Indonesia standard also lacks quite a few items that showed no variation such as items 10, 11, and 27. Thus, we believe that this measure of observed classroom quality may be more suitable than the original ECERS-R total score for our analysis of preschool quality and child development.

Second, we re-estimated Model 5 from our regression analyses above by using the Indonesia standard instead of the ECERS-R. We call this Model 6. It is our preferred model specification as it corrects for measurement error using instrumental variable and also uses a measure of observed classroom quality that is more locally relevant. The results of Model 6 are presented in Table 11.¹⁹

The results of Model 6 show that utilizing the Indonesia standard yields fairly consistent significant relationships between observed classroom quality and child development outcomes. For the social competence domain of the EDI, a one unit increase in the Indonesia standard is associated with a 0.121 unit increase in child development. Similarly, the size of the coefficient is 0.167 units for the language and cognitive development domain – the largest association for this model. For the physical health and well-being domain; a one unit increase in the Indonesia standard is associated with a 0.107 unit increase. Likewise for emotional maturity the estimated coefficient is 0.088 units.

Similar to previous estimations, Model 6 also shows mixed findings for teacher characteristics. Increasing the proportion of teachers with senior secondary or post-secondary education is associated with higher EDI in social competence and communication and general knowledge. The point estimates are fairly large relative to observed classroom quality. A 100% increase in teachers with post-secondary education is associated with a 1.078 to 1.188 unit increase in EDI. In contrast, increasing the mean years of teaching in a center is negatively associated with physical health and well-being and communication and general knowledge. While the coefficient is negative, the magnitude is fairly small relative to teacher's education, ranging from -0.015 to -0.045 units. Neither experience in ECED nor teacher training are significant predictors of child development. For structural characteristics, hours of operation per week is a positive predictor of EDI, with point estimates ranging from 0.027 (social competence) to 0.036 (language and cognitive development). Student-to-staff ratio is a significant positive predictor only for physical health and well-being (0.013 units) and the magnitude is very small compared to other predictors in the model.

The quality of Indonesian ECED services in context

The second analysis we conduct serves to tackle the question of how to use ECERS-R to

¹⁸ Such an approach necessarily limits our ability to make comparisons to other settings where ECERS-R has been applied.

¹⁹ Unstandardized regression coefficients of Model 6 are presented in Annex A (along with Models 1-5)

inform policy makers and the early childhood education community in Indonesia about the quality of preschool services in the country and how it compares with other parts of the world. First, when we examine the range of quality using ECERS-R, we find that not a single center in our sample score a 6 or 7 (excellent) on the total ECERS-R scale and the vast majority of services score 3 (minimal) or below (see Figure 2).

Based on these numbers alone, we might conclude that the quality of preschools in rural Indonesia is very low. But if we look across a number of other studies that have carried out an assessment of pre-school quality using ECERS-R, we find that services in rural Indonesia are not unique in their inability to score well on the ECERS-R (see Figure 3). Even services in Sweden or parts of Canada do not score above a 5 (good) on this scale on average. This suggests that the ECERS-R sets a high bar for excellence indeed.

In our textual analysis of comparing the national ECED standard to the items of the ECERS-R, we find that the definition of “minimal” is different between the ECERS-R and the Indonesia ECED Standard. Scoring of the ECERS-R begins by assessing whether all the requirements for a score of 1 are met; if they are, the rater assess whether all the indicators for a score of 3 are achieved. If only some of the indicators for a score of 3 are met, the center receives a score of 2. The indicators for a score of 3 on the ECERS-R are more exhaustive than what is listed in the Indonesia ECED Standard. This is not surprising given that the Indonesia ECED Standard sets minimal guidelines—they are meant to articulate the baseline requirement for preschool services. As a result, a center that meets the minimum requirements under the Indonesia ECED Standard receives a score of 2 on the ECERS-R scale rather than the “minimal” score of 3 as defined by the ECERS-R.

In Figure 4, we show the quality of preschool in Indonesia using two definitions of “minimal”—one using an ECERS-R score cutoff of 2 as in the Indonesia ECED Standard and one using an ECERS-R score cutoff of 3. The graph shows that using a context-specific definition of minimal standard presents a strikingly different picture than the ECERS-R definition. 81.0% of centers in rural Indonesia meet the national requirements for preschool quality while only 42.7% meet the ECERS-R minimum score. Given the resource-constrained environment in rural villages in Indonesia, we believe that using the Indonesia definition of minimal quality may be more useful to frame the conversation for national policy makers as well as the early childhood community in Indonesia. It is worth noting that the results in Figure 4 suggest that poor, rural communities are generally making efforts to meet the national standards. Despite these efforts, nearly 20% of early childhood services still need to achieve even a minimum requirement of quality as defined at the national level. Over 57% of centers fall short of meeting minimum requirements of quality at an internationally acceptable level. As the government seeks to expand access to preschool services more broadly it will need to increase its focus on enforcing its minimum service standards.

8. Discussion and Conclusion

Our analysis of early childhood education quality in Indonesia makes three important contributions to the literature. We find that:

- i. Classroom observations of ECED quality are subject to considerable measurement error. Not correcting for this leads to an underestimation of the association between quality and child development outcomes. Correcting for measurement error using an instrumental variables approach (with the first rater’s score as an instrument for the second rater’s score) allows us to improve the use of ECERS-R as a predictor of child development outcomes;

- ii. There is considerable variation in quality across types of preschool services within Indonesia. Assessed using ECERS-R, kindergartens run by the Ministry of Education and Culture (MoEC) and playgroups established under the MoEC's ECED project have noticeably higher quality than kindergartens run by the Ministry of Religious Affairs (MoRA) and playgroups not supported by the ECED project. When assessed using the Indonesian minimum service standards, 87 percent of project supported playgroups meet the government's standards compared to 79 percent of MoEC or MoRA kindergartens and 69 percent of non-project supported playgroups;²⁰ and
- iii. In countries with a national early childhood education standard, using a subset of the ECERS-R items that correspond with the national standard can provide researchers with an alternative measure of classroom quality that aligns closely to the particular context.

The estimated effect sizes of observed classroom quality for four out of five EDI domains (physical health, social competence, language and cognitive development, and communication and general knowledge) suggests that observed classroom quality is a modest, reliable predictor of children's developmental outcomes during early childhood in rural Indonesia. For the most complex model in this study (Model 6), a 1 standard deviation increase in classroom quality related to a 0.071 to 0.082 standard deviation increase in children's developmental outcomes. This range of effect sizes appears to be within the range of other empirically rigorous studies (i.e., those that include extensive control variables and/or use fixed effects) on classroom quality of early childhood programs both in the United States and in developing countries. For example, large-scale studies from the United States have effect sizes ranging from 0.04 (Duncan, 2003) to 0.18 (Peisner-Feinberg et al., 2001) on children's language and cognitive development. For socio-emotional skills, prior studies report effect sizes as small as 0.02 in the U.S. (Peisner-Feinberg et al., 2001) to as large as 0.13 in the U.K. (Sylva et al., 2006). Meanwhile, an East African study reports an effect size of 0.15 on children's cognitive skills (Malmberg et al., 2011). A study from Ecuador, which uses CLASS instead of ECERS-R, shows similar relationships between classroom quality and children's learning outcomes with effect sizes ranging from 0.06 to 0.13.

Compared to the few previous studies from developing countries that have examined the association between preschool quality and child development outcomes, our regression analyses introduced an extensive set of controls to adjust for potential biases that might lead to inconsistent estimates of observed classroom quality. Thus, our study confirms that even with these extensive controls (i.e., child, teacher, structural characteristics with district fixed effects), observed classroom quality is a significant and meaningful positive predictor of child development outcomes.

In addition to observed classroom quality, teacher's education is a significant predictor of children's social competence as well as communication and general knowledge. The effect size of teacher's education, particularly post-secondary education, is relatively large and ranges from 0.184-0.264 standard deviations in these two EDI domains. This suggests that the Government of Indonesia's policy to hire teachers with at least senior secondary education in early childhood education programs is supported by the results of our study.

²⁰ Any apparent jump in the relative quality of various kindergartens stems from the fact that Islamic kindergartens are bunched together at the lower end of the ECERS-R spectrum. While the ECERS-R uses a score of 3 as minimum, we interpret the Indonesia standard's minimum to be a score of 2 on the ECERS-R scale. As a result, a lot more Islamic kindergartens meet the Indonesia standards (looks more like the other types of centers) when the 58% of centers scoring a 2 are included in meeting the minimum quality level.

In contrast, mean years of teaching is negatively correlated with physical health and communication and general knowledge—although the effect sizes are relatively small, ranging from -0.056 to -0.117 standard deviations. In addition, experience in early childhood education and receiving teacher training yielded null results. For the case of teacher training, follow-up assessments of the training experience suggest that the intent of the model was inconsistently and infrequently followed. Often, in practice, training was didactic or overly theoretical. Follow-up on-site supervision has been expected but very rarely occurred.

It is also important to note, that teacher characteristics used in this study are averaged at the center-level, which means we cannot attribute student outcomes directly to their teachers. As a result, we interpret the small but negative coefficient on years of teaching as well as the null results of ECED experience and teacher training as follows: policies focused solely on hiring teachers with more experience and training will not be sufficient to improve children's development in rural Indonesia. Rather, policies must address the quality of professional development activities for teachers in order to ensure the effectiveness of early childhood education programs. Training, in particular, should be multifaceted: including not just group training but coaching, observations of quality programs, and regular teacher cluster meetings on meaningful topics.

The mixed results on teacher characteristics in our study is not surprising given that prior studies on teacher qualification have also shown mixed evidence in relation to children's developmental outcomes. The few studies that have successfully produced statistically significant results often employ a value-added model (Hanushek and Rivkin, 2006). A limitation of our teacher level data is that we cannot match the teacher observations with the student EDI. As a result, the measure of teacher characteristics used in this study is averaged at the center-level, which is a less precise measure of teacher quality than those derived from matched teacher-student datasets. Future studies of early childhood quality in developing countries could benefit from matched teacher-student data to improve the estimates of teacher quality and better understand what kinds of qualification of ECD teachers are particularly successful in producing high-level student outcomes.

Another explanation for the mixed results may be due to collinearity between measures of teacher qualifications and items on the ECERS-R that focus on teacher behavior in the classroom (e.g., Item 16 - Encouraging children to communicate). Although none of the ECERS-R items directly measure teacher's education, experience, and training, it is likely that those with higher levels of education, more experience, and better training tend to provide higher quality care as measured by the items on the ECERS-R, making it difficult to tease apart classroom quality and teacher qualifications. It is also plausible that some selection bias is taking place, whereby teachers with more education and/or experience choose to work in higher quality settings.

Finally, our analysis suggests that the quantity of exposure to early childhood education also matters. On average, children enrolled in more hours of early childhood programs in rural Indonesia scored higher on the EDI, when controlling for various child, teacher, and structural characteristics. Although previous studies examining the relationship between duration of preschool and child development outcomes have been mixed (Vandell, 2004), we find a positive association despite the fact that children are exposed to a low dose of early childhood education in rural Indonesia. On average, children in our sample were between the ages of 4 to 5, attending kindergartens for 2 to 3 hours a day for 6 days a week or in playgroups for 2 to 3 hours a day for 4 to 5 days per week. The maximum number of hours in a week was 24 hours, which was only found in 3 kindergartens in the sample. In contrast, studies that have found negative effects of

quantity of early care often focus on children who spend a substantial amount of time in care settings (i.e., 45 hours a week) over an extended period (Vandell, 2004). As Indonesia seeks to expand the range of early childhood interventions, such increases in dosage could also be achieved by incorporating children into community-level parenting programs which reinforce what is included in the preschool curriculum.

From a policy perspective, our study confirms the importance of investing in high-quality early childhood education programs. As Indonesia considers expanding early childhood education, there is a need to ensure that programs meet the national ECED standard if they are to have a meaningful impact on children's development.

Tables and Figures

Table 1: Types of early childhood programs in Indonesia

Type	Content	Responsible Ministry
Kindergarten (<i>Taman Kanak-kanak, TK</i>)	Pre-primary education	Ministry of Education and Culture
Islamic kindergarten (<i>Radhatul Athal, RA</i>)	Pre-primary education	Ministry of Religious Affairs
Playgroups (<i>Kelompok Bermain, KB</i>)	Early childhood education	Ministry of Education and Culture
Daycare (<i>Taman Penitipan Anak, TPA</i>)	Care service for children of working parents	Ministry of Social Welfare & Ministry of Education and Culture
Integrated health service unit (<i>Posyandu</i>)	Health care service for children and parenting information for mothers	Ministry of Health
Toddler family groups (<i>Bina Keluarga Balita, BKB</i>)	Parenting classes for mothers	National Family Planning Board

Note: This is a non-exhaustive list to show some of the most common types of early childhood programs in Indonesia.

Source: Hasan, Hyson and Chang, 2013

Table 2: Summary statistics of child-level variables by type of ECED

	Kindergarten		Islamic Kindergarten		Non-Project Playgroup		Project Playgroup		Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Child development outcomes										
EDI: Physical health & well-being	8.234	1.375	8.228	1.430	7.927	1.498	7.996	1.478	8.101	1.442
EDI: Social competence	6.927	1.724	6.812	1.627	6.755	1.683	6.634	1.679	6.778	1.698
EDI: Emotional maturity	6.532	1.433	6.274	1.245	6.099	1.316	6.320	1.336	6.373	1.373
EDI: Language & cognitive	7.229	2.456	7.240	2.534	6.285	2.565	6.189	2.653	6.696	2.607
EDI: Communication & general knowledge	6.188	2.114	6.193	1.995	6.033	2.030	5.773	2.020	6.002	2.065
Child characteristics										
Age (years)	5.124	0.723	5.010	0.948	4.407	1.014	4.331	1.155	4.707	1.043
Female (1=Yes)	0.504	0.500	0.520	0.500	0.513	0.500	0.500	0.500	0.505	0.500
Special needs (1=Yes)	0.038	0.192	0.056	0.229	0.023	0.151	0.037	0.190	0.038	0.190
Mother's highest education level:										
<i>Primary or less</i>	0.378	0.485	0.410	0.492	0.408	0.492	0.430	0.495	0.405	0.491
<i>Junior secondary</i>	0.277	0.448	0.267	0.443	0.271	0.444	0.263	0.441	0.270	0.444
<i>Senior secondary</i>	0.267	0.442	0.225	0.418	0.239	0.427	0.250	0.433	0.253	0.435
<i>Post secondary</i>	0.078	0.269	0.099	0.299	0.082	0.275	0.056	0.230	0.071	0.258

Note: EDI scores range from 1-10. All data are for children who are interviewed while enrolled in the type of center noted in the column heading.

Table 3: Summary statistics of center-level variables by type of ECED service

	Kindergarten (N=221)		Islamic kindergarten (N=50)		Non-project playgroup (N=70)		Project playgroup (N=236)		Total (N=578)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Observed classroom quality										
ECERS-R: Space & furnishing	2.915	1.223	2.636	0.993	2.522	1.075	3.235	1.251	2.974	1.223
ECERS-R: Personal care & routine	2.534	1.113	2.195	0.811	2.435	0.990	2.612	1.091	2.524	1.070
ECERS-R: Language-reasoning	3.762	1.593	3.023	1.131	3.204	1.401	3.518	1.215	3.531	1.404
ECERS-R: Activities	2.429	0.974	1.920	0.689	2.122	0.924	2.752	0.893	2.480	0.951
ECERS-R: Interactions	4.076	1.581	3.926	1.305	3.849	1.559	4.158	1.436	4.069	1.498
ECERS-R: Program structure	2.743	1.470	2.178	1.138	2.510	1.422	2.883	1.314	2.723	1.387
ECERS-R: Parents & staff	2.784	0.961	2.387	0.821	2.473	1.065	2.465	0.867	2.582	0.938
ECERS-R: Total (mean) score	3.035	1.049	2.609	0.725	2.731	0.995	3.089	0.911	2.983	0.974
(N=566 centers)										
Teacher characteristics										
Highest edu. level: Primary	0.000	0.000	0.000	0.000	0.014	0.084	0.006	0.051	0.004	0.044
Highest edu. level: Junior secondary	0.005	0.067	0.017	0.084	0.029	0.145	0.019	0.090	0.014	0.091
Highest edu. level: Senior secondary	0.347	0.402	0.502	0.393	0.684	0.392	0.742	0.342	0.563	0.417
Highest edu. level: Post secondary	0.648	0.403	0.481	0.402	0.273	0.388	0.233	0.328	0.419	0.418
Mean years of teaching	9.815	6.392	8.249	5.429	5.799	4.372	5.189	2.614	7.300	5.273
Teacher's prior experience is:										
ECED teacher experience	0.425	0.437	0.330	0.422	0.334	0.423	0.245	0.389	0.333	0.422
ECED teacher & non-teacher exp.	0.142	0.309	0.163	0.324	0.133	0.310	0.132	0.308	0.139	0.309
ECED non-teacher experience	0.171	0.333	0.152	0.269	0.173	0.316	0.318	0.403	0.229	0.363
No ECED experience	0.262	0.390	0.355	0.414	0.361	0.427	0.304	0.396	0.299	0.400
Teacher's training is:										
No training	0.164	0.317	0.323	0.398	0.144	0.287	0.117	0.255	0.157	0.303
Non-project training	0.834	0.317	0.677	0.398	0.856	0.287	0.259	0.347	0.592	0.433
100 hours of Project training	0.000	0.000	0.000	0.000	0.000	0.000	0.101	0.265	0.040	0.175
200 hours of Project training	0.002	0.022	0.000	0.000	0.000	0.000	0.522	0.412	0.210	0.366
Structural characteristics of center										
Student-to-staff ratio	12.162	5.788	9.419	5.152	9.182	4.821	10.214	4.383	10.769	5.199
Hours of operation per week	15.197	2.163	14.790	3.077	11.289	3.810	9.834	3.792	12.508	4.063

Note: N represents number of centers. ECERS-R scores are the averages of two observers who each rated the same centers at the same time using the stop-score rule. All teacher characteristics are averaged at the center-level and thus, continuous. Each variable is constructed by dividing the # of observed teachers in a center who have the particular characteristic by the total number of observed teachers in that center.

Table 4: Analysis of classroom quality, teacher, and structural characteristics on children's physical health & well-being

	DV: Physical health and well-being (Raw score)				
	no controls	w/ controls	w/ teacher	w/ teacher & structure	IV
	(1)	(2)	(3)	(4)	(5)
Observed classroom quality					
ECERS-R	0.107** (0.046)	0.073 (0.049)	0.079 (0.049)	0.083* (0.050)	0.107* (0.059)
Teacher characteristics					
% of teachers w/ sr. secondary edu.			0.275 (0.378)	0.297 (0.381)	0.283 (0.386)
% of teachers w/ post secondary edu.			0.272 (0.386)	0.285 (0.390)	0.278 (0.395)
Mean years of teaching			-0.013* (0.008)	-0.014* (0.008)	-0.015** (0.008)
% of teachers w/ no ECED experience			-0.048 (0.099)	-0.046 (0.099)	-0.057 (0.099)
% of teachers w/ no training			0.007 (0.133)	0.018 (0.134)	0.017 (0.134)
Structural characteristics					
Student-to-staff ratio				0.013* (0.007)	0.013* (0.007)
Hours of operation per week				0.008 (0.012)	0.008 (0.012)
Constant	7.774*** (0.142)	4.320*** (0.325)	4.075*** (0.497)	3.906*** (0.512)	3.860*** (0.522)
Observations	7,984	7,984	7,984	7,984	7,984
R-squared	0.005	0.153	0.155	0.157	0.155
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument					647.760
Control included:					
Child characteristics	N	Y	Y	Y	Y
Kindergarten dummy	N	Y	Y	Y	Y
District dummy	N	Y	Y	Y	Y

Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother's highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Model 5 is a two-stage least squares model to correct for measurement error in the ECERS-R. The F-statistic of the first stage is reported.

* p<0.1, ** p<0.05, *** p<0.01

Table 5: Analysis of classroom quality, teacher, and structural characteristics on children's social competence

	DV: Social competence (Raw score)				
	no controls	w/ controls	w/ teacher	w/ teacher & structure	IV
	(1)	(2)	(3)	(4)	(5)
Observed classroom quality					
ECERS-R	0.150*** (0.051)	0.096* (0.053)	0.083 (0.052)	0.070 (0.052)	0.111* (0.063)
Teacher characteristics					
% of teachers w/ sr. secondary edu.			0.825** (0.335)	0.826** (0.336)	0.808** (0.340)
% of teachers w/ post secondary edu.			1.103*** (0.347)	1.086*** (0.351)	1.073*** (0.356)
Mean years of teaching			-0.005 (0.009)	-0.006 (0.009)	-0.007 (0.009)
% of teachers w/ no ECED experience			0.089 (0.127)	0.067 (0.126)	0.055 (0.126)
% of teachers w/ no training			-0.125 (0.155)	-0.110 (0.156)	-0.111 (0.156)
Structural characteristics					
Student-to-staff ratio				0.001 (0.009)	0.002 (0.009)
Hours of operation per week				0.028* (0.016)	0.028* (0.015)
Constant	6.317*** (0.154)	3.084*** (0.420)	2.347*** (0.540)	2.257*** (0.556)	2.188*** (0.565)
Observations	7,957	7,957	7,957	7,957	7,957
R-squared	0.007	0.134	0.141	0.144	0.141
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument					646.386
Control included:					
Child characteristics	N	Y	Y	Y	Y
Kindergarten dummy	N	Y	Y	Y	Y
District dummy	N	Y	Y	Y	Y

Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother's highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Model 5 is a two-stage least squares model to correct for measurement error in the ECERS-R. The F-statistic of the first stage is reported.

* p<0.1, ** p<0.05, *** p<0.01

Table 6: Analysis of classroom quality, teacher, and structural characteristics on children’s emotional maturity

	DV: Emotional maturity (Raw score)				
	no controls	w/ controls	w/ teacher	w/ teacher & structure	IV
	(1)	(2)	(3)	(4)	(5)
Observed classroom quality					
ECERS-R	0.141*** (0.048)	0.072 (0.051)	0.063 (0.051)	0.063 (0.051)	0.072 (0.062)
Teacher characteristics					
% of teachers w/ sr. secondary edu.			0.191 (0.410)	0.191 (0.411)	0.183 (0.415)
% of teachers w/ post secondary edu.			0.271 (0.429)	0.271 (0.431)	0.268 (0.434)
Mean years of teaching			0.010 (0.008)	0.010 (0.008)	0.009 (0.008)
% of teachers w/ no ECED experience			0.064 (0.103)	0.063 (0.101)	0.055 (0.100)
% of teachers w/ no training			0.113 (0.129)	0.113 (0.129)	0.113 (0.129)
Structural characteristics					
Student-to-staff ratio				0.000 (0.007)	0.000 (0.007)
Hours of operation per week				0.001 (0.013)	0.002 (0.013)
Constant	5.944*** (0.137)	4.077*** (0.256)	3.817*** (0.490)	3.812*** (0.502)	3.792*** (0.504)
Observations	7,946	7,946	7,946	7,946	7,946
R-squared	0.010	0.116	0.118	0.118	0.117
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument					647.316
Control included:					
Child characteristics	N	Y	Y	Y	Y
Kindergarten dummy	N	Y	Y	Y	Y
District dummy	N	Y	Y	Y	Y

Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother’s highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Model 5 is a two-stage least squares model to correct for measurement error in the ECERS-R. The F-statistic of the first stage is reported.

* p<0.1, ** p<0.05, *** p<0.01

Table 7: Analysis of classroom quality, teacher, and structural characteristics on children’s language and cognitive development

	DV: Language & Cognitive (Raw score)				
	no controls	w/ controls	w/ teacher	w/ teacher & structure	IV
	(1)	(2)	(3)	(4)	(5)
Observed classroom quality					
ECERS-R	0.129 (0.079)	0.107 (0.080)	0.111 (0.080)	0.102 (0.081)	0.132 (0.096)
Teacher characteristics					
% of teachers w/ sr. secondary edu.			-0.135 (0.558)	-0.109 (0.550)	-0.127 (0.544)
% of teachers w/ post secondary edu.			-0.167 (0.580)	-0.170 (0.573)	-0.180 (0.568)
Mean years of teaching			-0.001 (0.012)	-0.003 (0.012)	-0.004 (0.012)
% of teachers w/ no ECED experience			0.274 (0.172)	0.255 (0.170)	0.241 (0.169)
% of teachers w/ no training			0.012 (0.220)	0.042 (0.217)	0.040 (0.217)
Structural characteristics					
Student-to-staff ratio				0.016 (0.011)	0.017 (0.011)
Hours of operation per week				0.039* (0.020)	0.039* (0.020)
Constant	6.323*** (0.248)	-1.128** (0.505)	-1.138 (0.765)	-1.437* (0.786)	-1.496* (0.785)
Observations	7,958	7,958	7,958	7,958	7,958
R-squared	0.002	0.266	0.268	0.270	0.269
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument					663.973
Control included:					
Child characteristics	N	Y	Y	Y	Y
Kindergarten dummy	N	Y	Y	Y	Y
District dummy	N	Y	Y	Y	Y

Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother’s highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Model 5 is a two-stage least squares model to correct for measurement error in the ECERS-R. The F-statistic of the first stage is reported.

* p<0.1, ** p<0.05, *** p<0.01

Table 8: Analysis of classroom quality, teacher, and structural characteristics on children's communication and general knowledge

	DV: Communication & General (Raw score)				
	no controls	w/ controls	w/ teacher	w/ teacher & structure	IV
	(1)	(2)	(3)	(4)	(5)
Observed classroom quality					
ECERS-R	0.151** (0.062)	0.073 (0.064)	0.085 (0.062)	0.080 (0.062)	0.120 (0.075)
Teacher characteristics					
% of teachers w/ sr. secondary edu.			0.935** (0.465)	0.926** (0.470)	0.908* (0.476)
% of teachers w/ post secondary edu.			1.206** (0.473)	1.196** (0.477)	1.183** (0.484)
Mean years of teaching			-0.044*** (0.011)	-0.043*** (0.012)	-0.045*** (0.012)
% of teachers w/ no ECED experience			-0.002 (0.139)	-0.009 (0.139)	-0.023 (0.139)
% of teachers w/ no training			0.074 (0.178)	0.073 (0.178)	0.072 (0.178)
Structural characteristics					
Student-to-staff ratio				-0.005 (0.012)	-0.004 (0.012)
Hours of operation per week				0.005 (0.018)	0.005 (0.018)
Constant	5.543*** (0.188)	0.976** (0.437)	0.041 (0.644)	0.084 (0.675)	0.015 (0.684)
Observations	7,981	7,981	7,981	7,981	7,981
R-squared	0.005	0.140	0.151	0.151	0.149
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument					647.668
Control included:					
Child characteristics	N	Y	Y	Y	Y
Kindergarten dummy	N	Y	Y	Y	Y
District dummy	N	Y	Y	Y	Y

Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother's highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Model 5 is a two-stage least squares model to correct for measurement error in the ECERS-R. The F-statistic of the first stage is reported.

* p<0.1, ** p<0.05, *** p<0.01

Table 9: Cross-walk of ECERS-R items and Indonesia ECED Standard

ECERS-R items	Is the ECERS-R item discussed in the Indonesia ECED Standard (2009)?	Citation from the Indonesia ECED Standard (2009) ¹
Space and Furnishing 1. Indoor space 2. Furniture for routine care, play and learning 3. Furnishings for relaxation and comfort 4. Room arrangement for play 5. Space for privacy 6. Child-related display 7. Space for gross motor play 8. Gross motor equipment	Yes Yes Yes	Section IV. A.2 Section IV.A.2 Section IV.A.2
Personal Care Routines 9. Greeting/departing 10. Meals/snacks 11. Nap/rest 12. Toileting/diapering 13. Health practices 14. Safety practices	(No, only for daycare) (No, only for daycare) Yes Yes Yes	Section IV.A.2 Section IV.A.2 Section IV.A.2 Section III.B.2 Section III.B.2
Language Reasoning 15. Books and pictures 16. Encouraging children to communicate 17. Using language to develop reasoning skills 18. Informal use of language	Yes Yes Yes	Section I.B.2.3 Section II.A.2.b.3 Section I.B.2-4
Activities 19. Fine motor 20. Art 21. Music/movement 22. Blocks 23. Sand/water 24. Dramatic play 25. Nature/science 26. Math/number 27. Use of TV, video and/or computers 28. Promoting acceptance of diversity	Yes Yes Yes Yes Yes Yes Yes	Section I.B.1-4 Section I.B.3 Section I.B.2-4 Section I.B.2-4 Section II.B.3-4 Section I.B.4 Section I.B.2-4
Interaction 29. Supervision of gross motor activities 30. General supervision of children 31. Discipline 32. Staff-child interactions 33. Interactions among children	Yes Yes Yes Yes	Section II.A.2.b.3 Section II.A.2.b.3 Section II.A.2.b.3 Section II.A.2.b.3,4
Program Structure 34. Schedule 35. Free play 36. Group time 37. Provisions for children with disabilities	Yes Yes Yes Yes	Section III.A.3 Section III.A Section III.A Section III.C.5
Parents and Staff 38. Provisions for parents 39. Provisions for personal needs of staff 40. Provisions for professional needs of staff 41. Staff interaction and cooperation 42. Supervision and evaluation of staff 43. Opportunities for professional growth	Yes Yes Yes Yes	Section II.A.2.b.2 Section II.A.2.b.4 Section III.C Section III.C.3

¹ Citation of Indonesia's national standards are coded as follows:

- | | |
|---|--|
| I. Standards for Development Achievement | II. Standards for Educators and Education Personnel |
| III. Standards for Content, Process, and Assessment | IV. Standards for Facility and Infrastructure, Management, and Financing |

Table 10: Mean and standard deviation of Indonesia standard

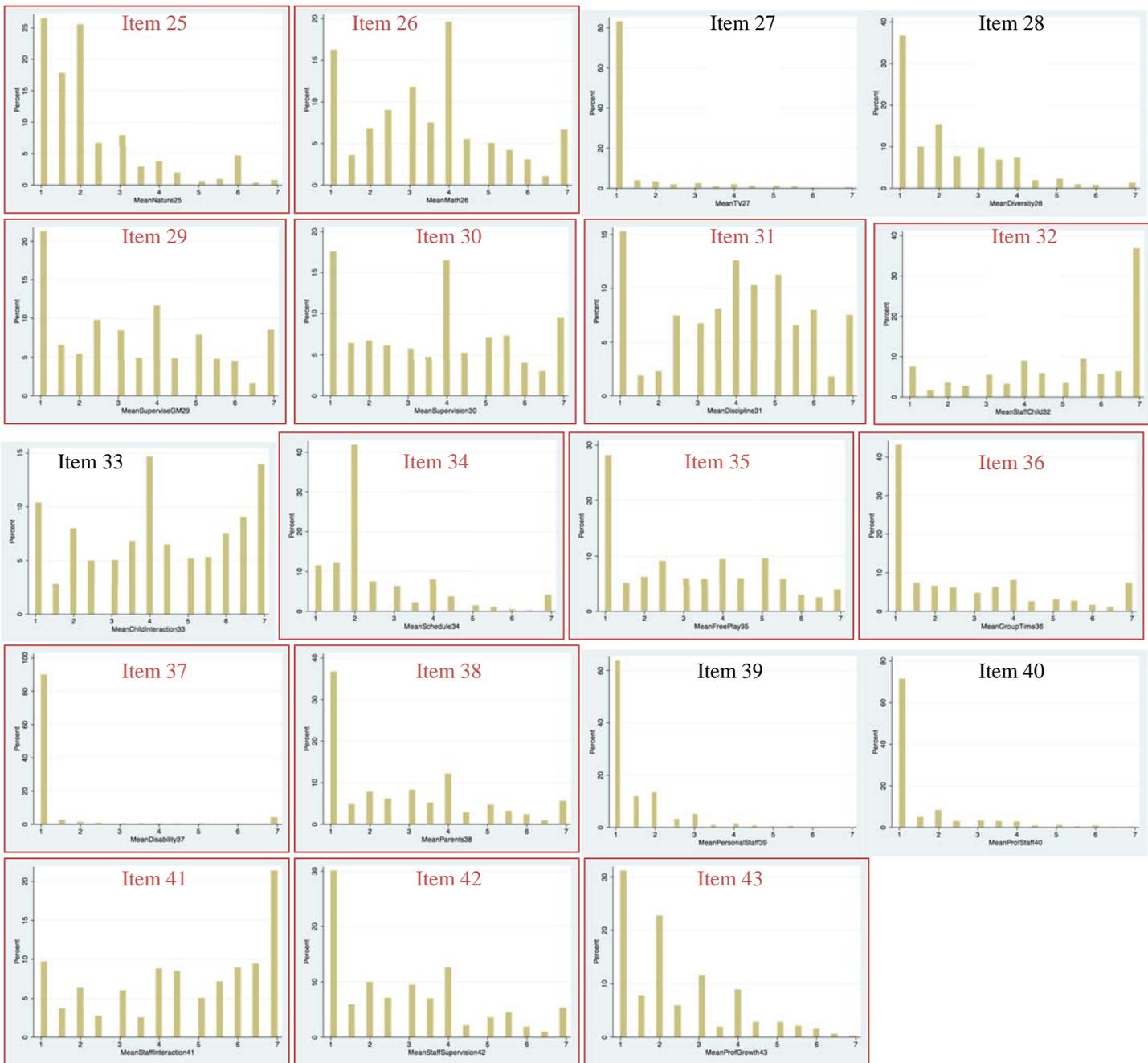
	Indonesia standard ¹	
	Mean	S.D.
Kindergarten	2.944	1.041
Islamic kindergarten	2.521	0.756
Non-project playgroup	2.659	1.007
Project playgroup	3.090	1.002
Overall	2.932	1.014

¹ The Indonesia ECED standard is an alternative measure of observed classroom quality that only includes the 28 common items between the ECERS-R and the national standards for ECED.

Figure 1: Histogram of raw responses on items in ECERS-R



Note: Red outline indicates items that align with the Indonesia ECED Standard



Note: Red outline indicates items that align with the Indonesia ECEC Standard.

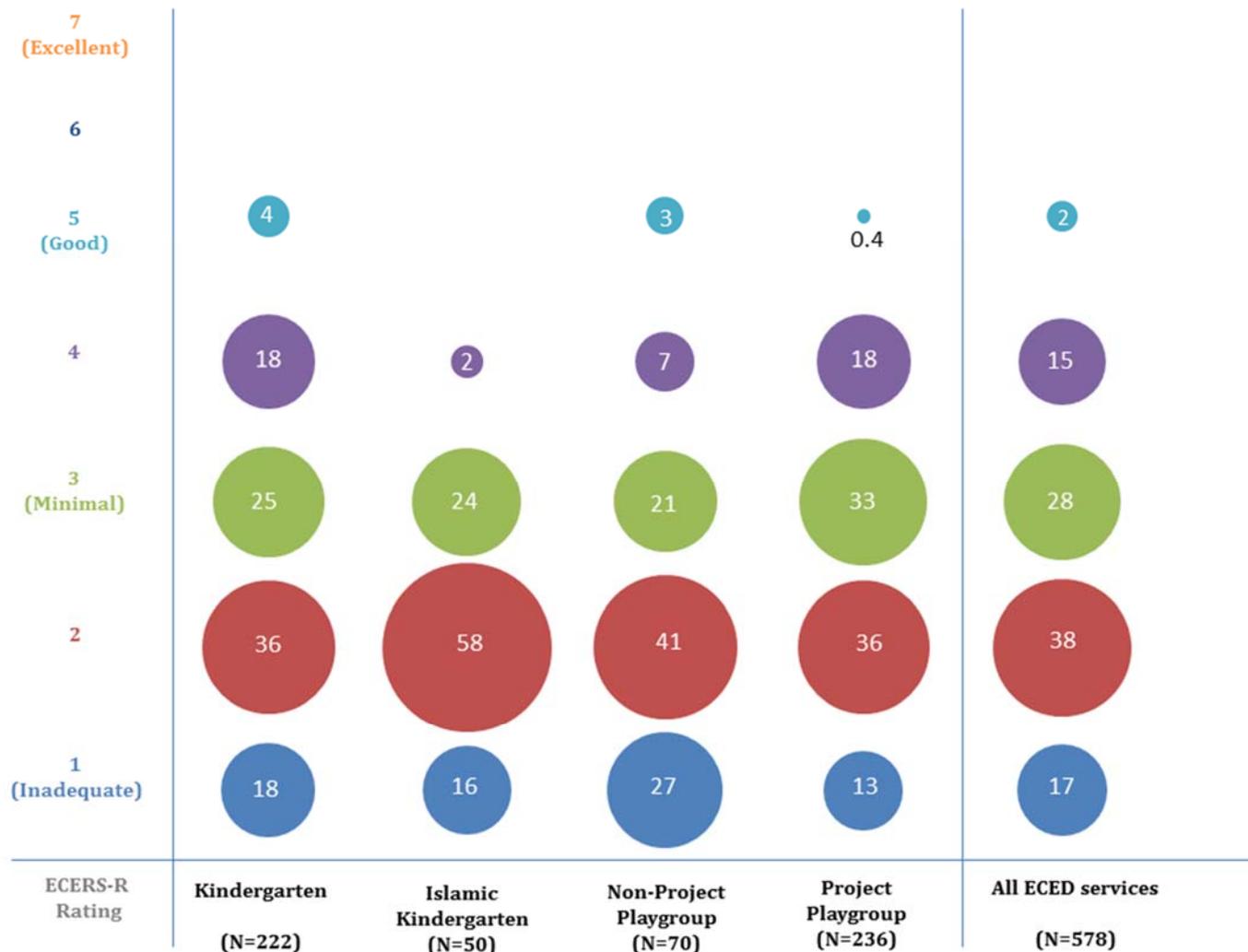
Table 11: Analysis of Indonesia standards on children's EDI (raw scores)

	DV: Physical health & well-being	DV: Social competence	DV: Emotional maturity	DV: Language & cognitive development	DV: Communication & general knowledge
Observed classroom quality					
Indonesia standard	0.107* (0.058)	0.121** (0.062)	0.088 (0.062)	0.167* (0.092)	0.133* (0.072)
Teacher characteristics					
% of teachers w/ sr. secondary edu.	0.288 (0.383)	0.813** (0.342)	0.184 (0.413)	-0.127 (0.538)	0.911* (0.481)
% of teachers w/ post secondary edu.	0.285 (0.392)	1.078*** (0.358)	0.269 (0.433)	-0.180 (0.563)	1.188** (0.489)
Mean years of teaching	-0.015** (0.008)	-0.008 (0.009)	0.009 (0.008)	-0.006 (0.012)	-0.045*** (0.012)
% of teachers w/ no ECED experience	-0.058 (0.099)	0.053 (0.125)	0.052 (0.100)	0.236 (0.170)	-0.025 (0.138)
% of teachers w/ no training	0.019 (0.133)	-0.110 (0.155)	0.114 (0.128)	0.041 (0.215)	0.073 (0.178)
Structural characteristics					
Student-to-staff ratio	0.013* (0.007)	0.002 (0.009)	0.001 (0.007)	0.017 (0.011)	-0.004 (0.012)
Hours of operation per week	0.007 (0.012)	0.027* (0.015)	0.000 (0.012)	0.036* (0.020)	0.004 (0.018)
Constant	3.876*** (0.518)	2.192*** (0.562)	3.781*** (0.502)	-1.524* (0.778)	0.014 (0.682)
Observations	7,984	7,957	7,946	7,958	7,981
R-squared	0.157	0.143	0.119	0.270	0.150
No. of clusters (centers)	565	565	565	565	565
First-stage F-statistic of excluded instrument	842.486	842.505	842.323	862.420	842.419
Control included:					
Child characteristics	Y	Y	Y	Y	Y
Kindergarten dummy	Y	Y	Y	Y	Y
District dummy	Y	Y	Y	Y	Y

* p<0.1, ** p<0.05, *** p<0.01

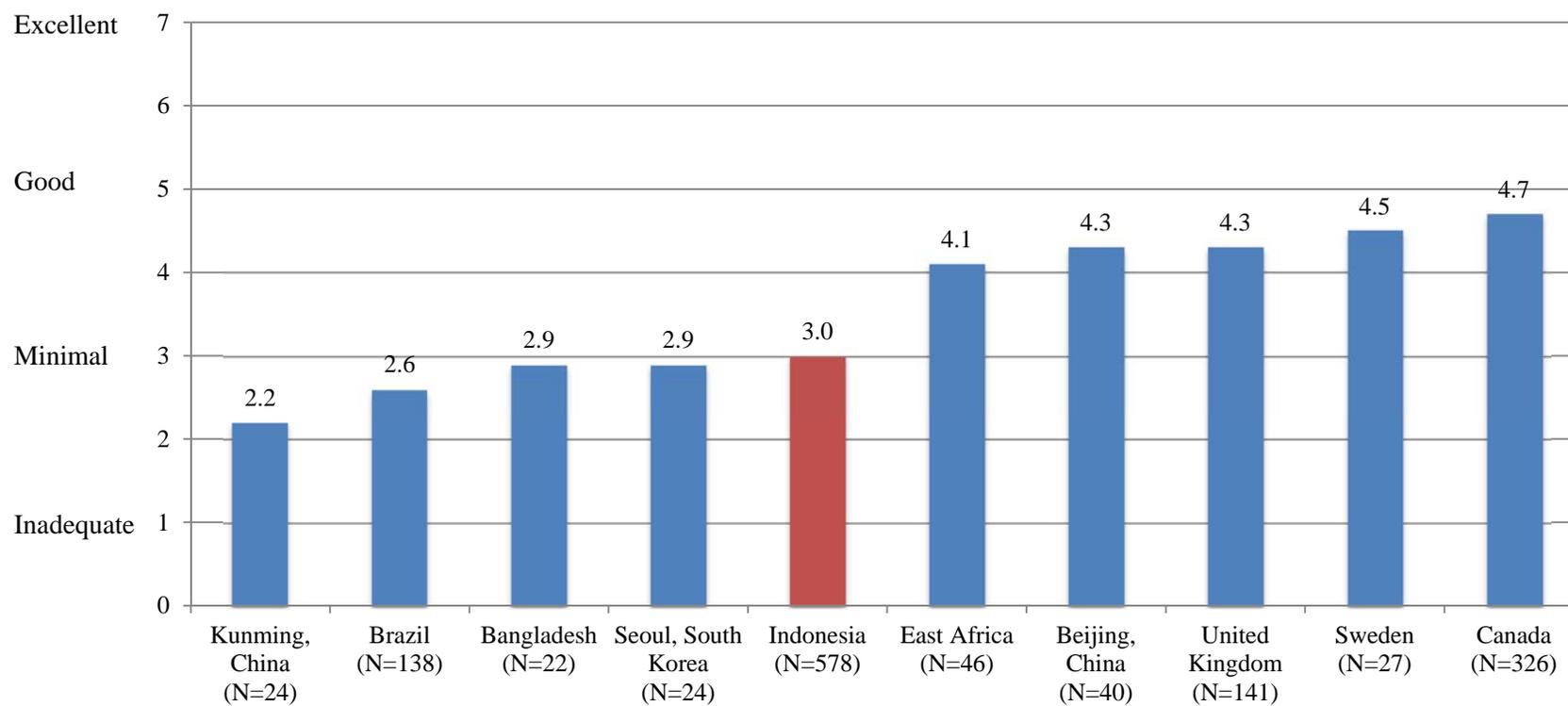
Note: Robust standard errors clustered at the center level in parentheses. Child characteristics include age, age-squared, gender, whether child has special needs (reported by teacher), and mother's highest education level. Kindergarten dummy is 1 for kindergartens or Islamic kindergartens and 0 for non-project playgroups or project playgroups. District dummies are included for the 8 districts where centers and children are located. Models 1-4 are OLS regressions. Models 5 and 6 are two-stage least squares model to correct for measurement error in the measure of center quality. The F-statistic of the first stage is reported.

Figure 2: Quality of early childhood education services in Indonesia, measured using ECERS-R



Note: Number inside bubble indicates percent distribution of ECERS-R score (adds up to 100% for each type of ECED service and for all ECED services). N denotes number of centers of each type.

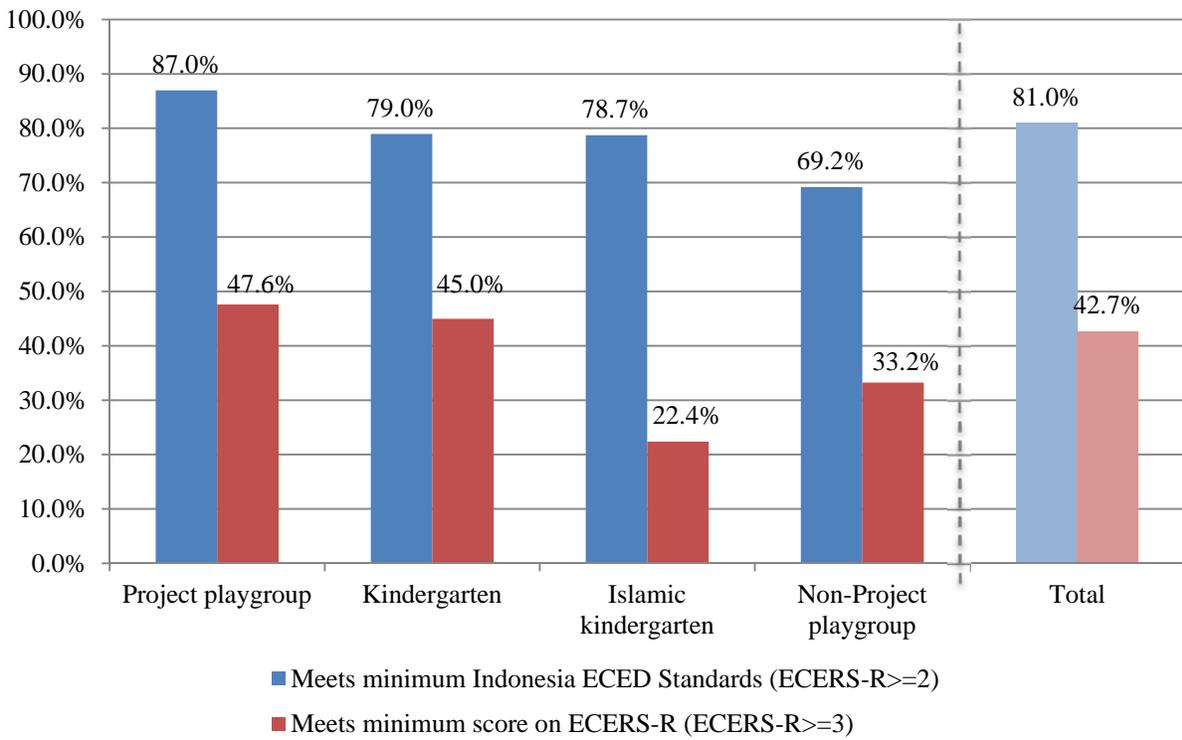
Figure 3: Average ECERS-R across various settings



N is the number of classrooms or ECED centers observed

Sources: Authors' calculations for Indonesia, Aboud (2006) for Bangladesh, Esposito et al. (2010) for Brazil, Goelman et al. (2006) for Canada, Liang et al (2013) for Kunming and Beijing, Malmberg et al. (2009) for East Africa, Sheridan et al. (2009) for Seoul and Sweden, Sylva et al. (2009) for United Kingdom.

Figure 4: Comparison of centers that meet the “minimal” threshold in the Indonesia ECED Standard and the ECERS-R



References

- About, F.E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, 21(1), 46-60. doi:10.1016/j.ecresq.2006.01.008
- Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. (2014). A helping hand? Teacher quality and learning outcomes in kindergarten. Washington, D.C.: *Inter-American Development Bank*.
- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5(3), 25-50.
- Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science*, 4(3), 136-148. doi:10.1207/S1532480XADS0403_3
- Brinkman, S., Gregory, T., Harris, J., Hart, B., Blackmore, S., and Janus, M. (2013). Associations between the Early Development Instrument at Age 5 and Reading and Numeracy Skills at Ages 8, 10 and 12: A Prospective Linked Data Study. *Child Indicators Research*, 6(4), 695-708. doi: 10.1007/s12187-013-9189-3
- Burchinal, M., Cryer, D., Clifford, R., and Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science*, 6(1), 2-11. doi:10.1207/S1532480XADS0601_01
- Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science*, 12(3), 140-153. doi:10.1080/10888690802199418
- Burchinal, M. R., Roberts, J. E., Riggins Jr, R., Zeisel, S. A., Neebe, E., and Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, 71(2), 339-357.
- Campbell, F. A. and Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, 65(2), 684-698. doi: 10.2307/1131410
- Clifford, R., Reszka, S., and Rossbach, H. (2010). Reliability and validity of the early childhood environment rating scale. *Working Paper*. Chapel Hill: FPG Child Development Institute, University of North Carolina.
- Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, 74(5), 1454-1475. doi:10.1111/1467-8624.00617
- Duncan, G. J. and Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 27(2), 109.
- Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., ... & Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten?. *Early Childhood Research Quarterly*, 21(2), 174-195.
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., ... & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child development*, 78(2), 558-580.
- Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., ... & Global Child Development Steering Group. (2011). "Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income

- countries." *The Lancet*, 378(9799): 1339-1353.
- Esposito, Y. L., Rosemberg, F., de Andrade, D.F., Gimenes, N., and Unbehaun, S. (2010). "An overview of the study: Early Childhood Education in Brazil. A qualitative and quantitative assessment". *Fundação Carlos Chagas, Brazilian Ministry of Education, and the Inter-American Development Bank*.
- Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2, 945-1017.
- Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., and LaGrange, A. (2006). Towards a predictive model of quality in Canadian child care centers. *Early Childhood Research Quarterly*, 21(3), 280-295.
- Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., and Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental psychology*, 49(1), 146. doi: 10.1037/a0027899
- Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies*. *The economic journal*, 113(485), F64-F98. doi: 10.1111/1468-0297.00099
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2, 1051-1078.
- Harms, T., Clifford, R. M., and Cryer, D. (2005). *Early Childhood Environment Rating Scale - Revised Edition*. New York: Teachers College Press.
- Hasan, A., Hyson, M., and Chang, M. (Eds.). (2013). *Early Childhood Education and Development in Poor Villages of Indonesia: Strong Foundations, Later Success*. Washington DC: World Bank.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27-50.
- Ip, P., Li, S. L., Rao, N., Ng, S. S. N., Lau, W. W. S., and Chow, C. B. (2013). Validation study of the Chinese Early Development Instrument (CEDI). *BMC Pediatrics*, 13(1), 146. doi: 10.1186/1471-2431-13-146
- Janus, M., Brinkman, S., and Duku, E. (2011). Validity and psychometric properties of the Early Development Instrument in Canada, Australia, United States and Jamaica. *Social Indicators Research*, 103(2), 283-297. doi: 10.1007/s11205-011-9846-1
- Janus, M., Brinkman, S., and Duku, E. (2007). *The EDI Handbook*. Ontario Canada: McMaster University Press.
- Janus, M. and Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science*, 39, 1-22.
- Liang, X., Zhang, Y., and Fu, Y. (2013). *Early childhood education in Yunnan: Challenge and opportunities*. Washington, D.C.: World Bank Group.
- Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? . *Economics of Education Review*, 26(1), 33-51. doi: 10.1016/j.econedurev.2005.09.008
- Malmberg, K. A., 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. doi: 10.1016/j.ecresq.2010.04.003

- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... and Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.
- Montes, G., Hightower, A. D., Brugger, L., and Moustafa, E. (2005). Quality child care and socio-emotional risk factors: No evidence of diminishing returns for urban children. *Early Childhood Research Quarterly, 20*(3), 361-372. doi: 10.1016/j.ecresq.2005.07.006
- National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71*(3), 960-980.
- National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten?. *Child Development, 74*(4), 976-1005.
- National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. (2005). Early Child Care and Children's Development in the Primary Grades: Follow-up Results from the NICHD Study of Early Child Care. *American Educational Research Journal, 42*(3), 537-570. doi: 10.2307/3700462
- Peisner-Feinberg, E. S. and Burchinal, M. R. (1997). Relations Between Preschool Children's Child-Care Experiences and Concurrent Development: The Cost, Quality, and Outcomes Study. *Merrill-Palmer Quarterly, 43*(3), 451-477. doi: 10.2307/23093333
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., and Yazejian, N. (2001). The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade. *Child Development, 72*(5), 1534-1553. doi: 10.1111/1467-8624.00364
- Phillips, D. A., Mekos, D., Scarr, S., McCartney, K., and Abbott-Shim, M. (2001). Within and beyond the classroom door: assessing quality in child care centers. *Early Childhood Research Quarterly, 15*(4), 475-496. doi: 10.1016/S0885-2006(01)00077-1
- Pradhan, M., Brinkman, S. A., Beatty, A., Maika, A., Satriawan, E., de Ree, J., and Hasan, A. (2013). Evaluating a community-based early childhood education and development program in Indonesia: study protocol for a pragmatic cluster randomized controlled trial with supplementary matched control group. *Trials, 14*(1), 259. doi:10.1186/1745-6215-14-259
- Rao, N. (2010). Preschool quality and the development of children from economically disadvantaged families in India. *Early Education and Development, 21*(2), 167-185. doi:10.1080/10409281003635770
- Sheridan, S., Giota, J., Han, Y. M., and Kwon, J. Y. (2009). A cross-cultural study of preschool quality in South Korea and Sweden: ECERS evaluations. *Early Childhood Research Quarterly, 24*(2), 142-156. doi:10.1016/j.ecresq.2009.03.004
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*(1), 76-92. doi: 10.1016/j.ecresq.2006.01.003
- Vandell, D. L., & Wolfe, B. (2000). *Child Care Quality: Does It Matter and Does It Need to Be Improved?* Washington, D.C.: Department of Health and Human Services.
- Vandell, D. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly, 50*(3), 387-414.
- Zachrisson, H., Dearing, E., Lekhai, R., Toppelberg, C. (2013). Little Evidence That Time in Child Care Causes Externalizing Problems During Early Childhood in Norway. *Child Development, 84*(4), 1152-1170.

Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer, L., and Margie, N. G. (2006). Child Outcome Measures in the Study of Child Care Quality. *Evaluation Review*, 30(5), 577-610. doi: 10.1177/0193841x06291529

Annex

A1. Results of Model 4 using each ECERS-R subscales for each EDI domain

	DV: Physical health & wellbeing	DV: Social competence	DV: Emotional maturity	DV: Language & cognitive	DV: Communication & general knowledge
Space & furnishing	0.096** (0.038)	0.089** (0.043)	0.041 (0.043)	0.146** (0.062)	0.080* (0.047)
Personal care	0.015 (0.045)	-0.013 (0.047)	0.015 (0.045)	0.033 (0.069)	-0.010 (0.057)
Language reasoning	0.030 (0.031)	0.016 (0.034)	0.020 (0.029)	0.035 (0.053)	0.016 (0.044)
Activities	0.080* (0.046)	0.102** (0.052)	0.126*** (0.048)	0.130* (0.077)	0.048 (0.061)
Interactions	0.051* (0.028)	0.040 (0.030)	0.018 (0.026)	0.016 (0.045)	0.078** (0.037)
Program structure	0.008 (0.034)	0.004 (0.037)	0.023 (0.036)	0.031 (0.054)	0.033 (0.043)
Parents & staff	0.127** (0.055)	0.127** (0.052)	0.107** (0.047)	0.128 (0.078)	0.099 (0.071)
Total ECERS-R	0.083* (0.050)	0.070 (0.052)	0.063 (0.051)	0.102 (0.081)	0.080 (0.062)

***p<.01; **p<.05; *p<.1

Note: Each cell is a result of a separate regression

In our study, we focus on ECERS-R as a holistic measure of center quality by using the total ECERS-R score (consistent with the use of ECERS-R in most existing studies). In the annex, we also provide evidence of how the different subscales of ECERS-R relate to child development outcomes. In Table A1, we report on the relationship between center quality and child development outcomes in Model 4 using each ECERS-R domain as the main independent variable. As a result, each cell is the output of a separate regression. For reference, we also include the coefficients of the total ECERS-R in the last row (same as Model 4 from Tables 4-8).

For the physical health and wellbeing domain, we find statistically significant positive correlations with four of the ECERS-R subscales. The largest positive correlations are found with parents & staff (0.127 units) and space & furnishing (0.096 units). We also find statistically significant positive correlations with activities (0.080 units) and interaction (0.051 units)—but these are much smaller in magnitude.

None of the other EDI domains had statistically significant correlations with the total ECERS-R score. However, we find some subscales of the ECERS-R to be significant positive predictors of

EDI. For social competence, parents & staff (0.127 units), activities (0.102 units), and space & furnishing (0.089 units) are statistically significant. For emotional maturity, activities (0.126 units) and parents & staff (0.107 units) are significant positive predictors. The language and cognitive development domain yields similar results with positive associations with space & furnishing (0.146 units) and activities (0.130 units). Finally, for communications and general knowledge, space & furnishing (0.080 units) and interaction (0.078 units) are significant positive predictors.

Overall, space & furnishing and activities are positively correlated with four out of five EDI domains while controlling for teacher and other structural characteristics (as specified in Model 4). While this would suggest that the provision of adequate materials and activities for children are important components of center quality, we suspect this also reflects the ease with which raters could discern these aspects of center quality.