RIETI Discussion Paper Series 24-E-058

Analyzing Childcare Quality: Impacts on child development and parental mental health, and effectiveness of professional development

FUJISAWA, Keiko K.
Keio University

FUKAI, Taiyo
Gakushuin University

LE, Chien Quang
Keio University

NAKAMURO, Makiko
RIETI

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 24-E-058
May 2024

Analyzing Childcare Quality: Impacts on Child Development and Parental Mental Health, and Effectiveness of Professional Development [1]

Keiko K. Fujisawa（Keio University）・Taiyo Fukai（Gakushuin University）・Chien Le Quang（Keio University）・Makiko Nakamuro（Keio University）

Abstract

This study examined the impact of childcare quality on child development and parents and the effectiveness of the professional development program in improving childcare quality. In Study 1, the childcare quality of 15 classes of 5-year-old children in all 14 licensed childcare centers under the jurisdiction of one municipality in Tokyo was assessed using the Early Childhood Environment Rating Scale, 3rd edition (ECERS-3; Harms et al., 2015, Translated by Uzuhashi, 2016). Data were obtained on each child's developmental status and problem behaviors at five years of age, academic achievement and problem behaviors at the second grade, and parental status at five years of age, including parents' feelings toward their children, nurturing behaviors, and mental health. The results of the analysis showed that the quality of childcare in the 5-year-old class positively affected parents' positive feelings toward their children, parents' good mental health, and academic achievement after school entry. In Study 2, 20 classes of 5-year-old children at licensed childcare centers and kindergartens in the same municipality were randomly assigned to the intervention group and the control group in a different year from Study 1, and a short program of professional development was conducted with the center directors and the caregivers in charge of the intervention group in the form of providing detailed feedback on the results of the ECERS-3. The results showed that scores on the ECERS-3 increased in the intervention group compared to the control group. More than half of the targeted indicators showed improvement.

Keywords: early childhood education and care, childcare centers, childcare quality, intervention, Early Childhood Environment Rating Scale, 3rd edition

JEL classification: J13, J18, I20, I28

[1]This study is conducted as a part of the project “Implementing Evidence-Based Policy Making in Japan” undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at the RIETI DP seminar for the paper. We would like to thank participants of the RIETI DP Seminar for their helpful comments. This study was approved by the Keio University SFC Ethics Committee for Experiments and Research (Reception No. 290).

The increasing participation of women in the labor market has increased the need for nonparental childcare, leading to a rise in childcare services worldwide. While the provision of universal access to early childhood education and care (ECEC) is a major policy issue, the effects of access to ECEC have been examined. Many studies have reported the mid- and long-term impacts of ECEC on various outcomes, such as parental employment and child outcomes (Burchinal et al., 2022). Findings emphasize that universal access to ECEC particularly benefits socioeconomically disadvantaged children and parents (Blau, 2021; Havnes & Mogstad, 2015; Yamaguchi et al., 2018).

ECEC quality and access may matter in the context of its universal access. Whitaker et al. (2023) found that the effect of ECEC has decreased in recent years, which might be because preschool programs have become less age-appropriate for preschool-age children, underscoring the importance of examining the effects of ECEC quality. Existing research suggests that ECEC quality is associated with a child's cognitive ability only if it is above a certain level (Vandell et al., 2010); in contrast, low-quality childcare harms child outcomes (Japel et al., 2005; Fort et al., 2020). While previous research could not fully rule out selection bias arising from possible sorting of children to high- or low-quality care, research on the causal relationship between ECEC quality and outcomes remains scarce. If a causal relationship exists, how to improve ECEC quality must be clarified. There is growing consensus (OECD, 2019) that professional development (PD) for ECEC teachers enhances ECEC quality. PD is in-service training for ECEC teachers, encompassing learning opportunities to facilitate the acquisition of professional knowledge, skills, and disposition in order to improve teaching and child outcomes (Egert et al., 2020). Nevertheless, there are no conclusions on effective PD content and delivery for ECEC teachers.

Causal relationship between ECEC quality and child outcomes

For several reasons, the association between ECEC quality and child outcomes should be studied in the context of universal ECEC access. First, studies that examined the effects of ECEC quality focused on targeted programs for disadvantaged children, such as Perry Preschool Program and Head Start, thus, these findings are not generalizable (Blau, 2021). Second, selection bias may confound findings of past studies because participation in ECEC is not random. In other words, the results of these studies are interpreted as correlational; thus, the causal relationship between ECEC quality and outcomes is inferred but not concluded (Perlman et al., 2016). One exception to this can be found in Araujo et al.'s (2019) study. Third, findings regarding the medium and longer-term associations between ECEC quality and child outcomes are inconsistent (e.g., Guerrero-Rosada et al., 2021; Oppermann et al., 2023; Peisner-Feinberg et al., 2001; Sylva et al., 2011).

PD for ECEC teachers to improve ECEC quality

Despite greater clarity on the definition and framework of PD, findings regarding the design and effectiveness of PD programs are highly heterogeneous (Brunsek et al., 2020). The scale-based PD approach, wherein training content is aligned with observation-based quality rating and in-service providers offer scale-based feedback to teachers (Egert et al., 2018; 2020), will likely be a highly effective method to enhance ECEC quality and children's development. Given that PD programs that align with child outcomes are considered the most effective, PD programs that respond to the observation measure may also be effective since it is used for examining the association between ECEC quality and child outcomes. Further, Blau (2000) showed that teacher training is an important determinant of ECEC quality measured by the Early Childhood Environment Scale (Harms, Clifford, & Cryer, 1998). Notably, the methodological

concern that scale-based PD might yield bias known as “training to the test” has not been confirmed (Egert et al., 2018). PD programs that include coaching are more effective than workshop-like programs, which are interpreted as the coaching includes more individualized attention and feedback tailored to meet the needs of ECEC teachers (Brunsek et al., 2020; Egert et al., 2020).

Participants are more likely to complete short-time programs and receive the intended content. Further, because other caregivers must supplement children’s care while caregivers participate in the PD program, short programs are less burdensome for childcare settings, where staffing shortages and long working hours are a challenge. Egert et al. (2018) reported a curvilinear relationship between the length and effectiveness of PD programs, with those of 45–60 hours being most effective at improving quality. However, no consistent conclusions have been reached regarding the impact of PD programs’ duration considering participant attrition (Brunsek et al., 2020). In this study, we developed a new PD program characterized by a scale-based approach, a short duration, and a coaching style that is tailored to fulfill the requirements of individual center directors and caregivers in a target class.

Unique context and characteristics of Japanese ECEC

Research on ECEC quality and PD outcomes has been predominantly undertaken in Europe and North America. Considering the numerous contextual factors, such as regulations for the provided quality and teacher qualification, findings concerning ECEC quality and PD in different cultural contexts can contribute to existing knowledge. The Japanese ECEC context is unique and provides a suitable environment for investigating the causal impact of and improving ECEC quality.

In Japan, universal ECEC for preschool-age children was nearly achieved as early as 1980s. The Japanese government subsidizes ECEC facilities, comprising kindergartens and childcare centers, with childcare centers accounting for high enrollment rates (Children and Family Agency, Japan, 2023). Childcare centers are available for children as young as 0 years old, and kindergartens are for children as young as 3 years old. Standard childcare centers in Japan operate for 11 hours a day. The legally regulated educational hours for kindergartens are 4 hours, but many are open before 9:00 a.m. and beyond 5:00 p.m. (Ministry of Education, Culture, Sports, Science and Technology, Japan, 2020). The child–teacher ratio by age and facility is legally regulated. Teacher qualifications are also regulated at the national level, and their academic backgrounds are homogenous (mostly junior college; Tokyo Metropolitan Government, 2022). While laws regulate the structure quality, such as classroom size and equipment, national regulations on process quality are relatively relaxed in Japan (see Ministry of Health, Labour and Welfare, Japan, 2018).

ECEC in Japan has historically emphasized child-centered educational practices. Caregivers have been expected to promote and assist children's spontaneous activities, respecting child-oriented free play rather than caregiver-driven activities. The caregiver's role is emphasized in the specific teaching plans based on understanding each stage of children's development to enable children to lead a life that is age-appropriate for early childhood and to gain the necessary experience and knowledge comprehensively (Ochanomizu University Center for Women's Education and Development, 2006). ECEC facilities also teach children to achieve personal independence in daily tasks, such as eating, dressing, and using toilets. These characteristics of Japanese ECEC contrast with ECEC in North America, where the emphasis is on promoting academic learning with an upward trend in teacher-directed instruction,

particularly in full-day classrooms (Markowitz & Ansari, 2020). Further, holistic support for parents has historically been an essential part of caregivers' work, especially in childcare centers (Ochanomizu University Center for Women's Education and Development, 2006). This anticipates the potential impact of ECEC quality on parents as well as children.

Fujisawa et al. (2023a) evaluated ECEC quality of accredited childcare centers in Japan using the Early Childhood Environment Rating Scale, 3rd edition (ECERS-3) (Harms, et al., 2015), which was the first published study of ECEC quality based on observation of trained assessors. The researchers found differences in ECEC quality in the U.S. and Japan and discussed the findings regarding the influence of the standard guidelines for accredited childcare centers. However, their study did not examine the relation between ECEC quality and outcomes.

Aims

This paper comprises Study 1 and Study 2. Study 1 examines whether high ECEC quality impact on better outcomes, and Study 2 determines how to improve childcare quality. Our study contributes to the literature in several ways. First, to our best knowledge, this is the first study to examine the causal relationship between childcare quality and short- and mid-term outcomes using longitudinal data. We collected data from a city in Japan and analyzed a representative population in the context of universal access. Second, we originally developed a PD program for this study to improve ECEC quality and evaluate the program's effectiveness by running a clustered randomized controlled trial design. The PD program involves experts who provide feedback to childcare providers for an hour at a time, which is very concise. Finally, our study based on Japan's ECEC context contributes to the literature that contains scarce research from Asian countries.

General Method

Overview

Data were collected in a Tokyo metropolitan area city. Collaborating with the city's school board of education, we invited all kindergartens (N = 4) and accredited childcare centers (N = 14) with a 5-year-old class in the city to participate. All excluding one kindergarten agreed to participate. The childcare center capacity in the target city was considered to meet the childcare demand as no child in the city was in the waiting list, indicating that universal access was achieved.

Study 1 included a longitudinal design, and only childcare centers were targeted. Fourteen childcare centers with 5-year-old classes (15 classes in total) participated in Study 1. Children and parents in the 5-year-old classes in the fiscal year (FY) 2020 participated in the survey on child development, behavioral problems, and parental socioeconomic status (boys: N = 107; girls: N = 93; unknown: N = 4). The rate of parental consent was 76.1% (204 of 268). The children were followed up until they reached Grade 2 in FY2022 and were examined for academic performance and behavioral problems (follow-up rate was 79.90%; boys: N = 82; girls: N = 77; unknown: N = 4). Study 1 assessed the 5-year-old class childcare quality and the effect of that quality in FY2020. Although not the focus of the analysis in Study 1, the quality of childcare at the same childcare centers was also evaluated in FY2021 and FY2022.

Study 2 targeted all childcare facilities with 5-year-old classes in FY2022. Three kindergartens and 14 childcare centers were randomly assigned to the intervention and control groups, and the PD program was implemented in the intervention group. Changes in childcare quality before and after intervention were analyzed.

Childcare quality

We used ECERS-3 (Harms et al., 2015) to assess ECEC quality for Studies 1 and 2. ECERS-3, based on a constructive theory of cognitive development, is an internationally comparative instrument and is widely used in more than 20 countries (Burchinal, 2018), including Japan (Fujisawa et al., 2023a). It assesses comprehensive childcare quality in six subscales: space and furnishings, personal care routines, language and literacy, learning activities, interaction, and program structure*.

ECERS-3 includes 461 indicators aggregated into 35 items and 6 subscales. Indicators are categorized into levels 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). When the criterion for each indicator is met, it is rated 1; otherwise, it is rated 0. Each item contains ten or more indicators, with scores ranging between 1 (inadequate) and 7 (excellent) (Harms et al., 2015). The subscale score is the mean of the items scores included in the subscale. The total score is the mean of the scores for all items.

In observation, two or three trained assessors visited the centers and scored 461 indicators independently. Observation time was from 9:00 a.m. to 12:30 (3.5 hours). Details regarding the

---

* In Japan, the Nursery Center Childcare Guidelines and the Course of Study for Kindergarten provide guidelines for ECEC, while they are comprehensive in nature and do not provide as detailed regulations as ECERS. However, the consistency between ECERS-R, an earlier version of ECERS-3, and these guidelines has been confirmed (Association of Private Kindergartens of Osaka, Japan, 2014), and the essential content of ECEC that makes up ECERS is not considered to be at odds with Japanese ECEC policy.

assessors' qualifications and training procedure were the same as described by Fujisawa et al. (2023a). The mean (SD) agreement rate between assessors was 87.15% (5.12). Any disagreement was agreed upon through discussion following observation.

Study 1

Method

Measures

Children's development before school entry

We asked primary caregivers to measure children's developmental age using the "KIDS Infant Development Scale" (KIDS), developed and standardized for Japanese children (Miyake et al., 1990); this scale has been validated (Hashimoto, 2013) and used in large-scale surveys for children in childcare centers (e.g., Sato et al., 2023). The primary caregivers rated 133 items "yes" or "no" based on each child's capability in eight domains: physical motor, manipulation, receptive language, expressive language, language concepts, social relationships with children, social relationships with adults, and discipline. The developmental age in months (DA) was calculated based on the number of items achieved. The developmental quotient (DQ) is calculated as follows: the DA is divided by the chronological age in months and multiplied by 100. The DQ was used as the KIDS score in subsequent analyses.

Parental feelings, discipline, and mental health before school entry

We collected information on parental feelings and mental health in the Parent Feelings Questionnaire with seven questions and the Parental Discipline Interview with nine questions using a five-point Likert scale ranging from "very often" to "never" (Deater-Deckard, 2000). We also included five questions from the World Health Organization's Five Well-Being Index to measure parents' mental health. For these three parental outcomes, we sum the relevant

responses. The parent questionnaire included children's demographic characteristics, such as gender, birth month, and birth weight, and family information, such as parental income and educational background.

Behavioral problems before and after school entry

Parents completed the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) in the 5-year-old class and Grade 2. The SDQ captures behavioral problems with a three-point Likert scale and has five components: (i) hyperactivity or inattention, (ii) conduct problems, (iii) emotional symptoms, (iv) peer problems, and (v) prosocial behavior. The total difficulties score, calculated by the sum of scores on (i), (ii), (iii), and (iv), captured behavioral problems.

Academic performance after school entry

All Grade 2 students from public schools in the city take the standardized achievement tests for reading and math. All test scores are transformed into z-scores for ease of interpretation, with mean 0 and SD 1. Using the questionnaire survey conducted along with the achievement tests, we also measured students' behavioral outcomes, such as hours studying at home.

Childcare quality

ECEC quality in the 5-year-old class was evaluated using ECERS-3 from June to August 2020. Note that the Japanese academic year begins in April.

Statistical analysis

Using the variation in childcare quality across class, this study estimates the effect of childcare quality on children's outcomes before and after school entry as follows.

$$y_{ic} = \beta_0 + \beta_1 ECERS_c + \beta_2 X_{ic} + \varepsilon_{ic},$$

where y is the outcome of child i in class c; $\boldsymbol{X}$ is a set of controls, such as the demographic characteristics of children and their parents; and ε_{ic} denotes the error term.

Results and Discussion

Descriptive statistics

Table 1 presents the descriptive statistics of child and parent outcomes, demographic characteristics, and ECERS-3. The average KIDS score was above 100, meaning the children showed better-than-average developmental at age 5. Children's behavioral problems at age 5 years and at Grade 2, as assessed by the SDQ's total difficulties score, were lower than those of the larger sample size studies (Iida et al., 2014; Moriwaki & Kamio, 2014). ECERS-3 score was also higher than in a previous study (Fujisawa et al., 2022a) conducted in other Japanese city (3.41 with 0.69 SD; see Appendix Table A1 for the subscale descriptive statics). The internal consistency of the variables was moderate to high, except for that of some ECERS-3 subscale scores that included only three or four items.

Table 1

Descriptive statistics of child and parent outcomes, and demographic characteristics

	Mean	SD	α	N
KIDS score at age 5 (before school entry)	106.787	16.097	0.958	198
Behavioral problems at age 5	29.057	4.952	0.745	192
Parental feeling at age 5				
Positive feeling	13.431	1.627	0.629	195
Negative feeling	11.635	3.466	0.812	192
Parental discipline at age 5	19.874	3.118	0.567	190
Parental mental health at age 5	16.079	5.013	0.884	190
Demographic characteristics at age 5				
Gender (female = 1)	0.465	0.500		200
Birth weight (> 2,500g)	0.905	0.294		200
Household income (thousand JPY)	608.840	288.683		181
Father's education (in years)	13.713	2.221		171
Mother's education (in years)	13.521	2.010		190
Academic performance at Grade 2				

Math (0–100)	77.749	17.275		163
Reading (0–100)	68.596	16.992		163
Hours of studying at home	0.778	0.471		117
Behavioral problems at Grade 2	28.530	4.444	0.727	117
ECERS-3 (FY2020)				
Total	4.605	0.351	0.699	15
Space and furnishings	4.571	0.833	0.481	15
Personal care routines	5.600	0.944	0.230	15
Language and literacy	4.170	0.948	0.445	15
Learning activities	3.187	0.484	0.697	15
Interaction	6.500	0.502	0.746	15
Program structure	5.778	0.851	0.271	15

Examining the parental selection issue

Conditional on children's and parents' observable characteristics, the causal relationship between ECEC quality and outcomes requires that ECEC quality is randomly assigned. Typical criticism is that highly educated parents opt for high-quality childcare (Fort et al 2020). The positive selection of childcare based on quality will generally lead us to overestimate the underlying relationship between childcare quality and subsequent outcomes.

It is relevant to discuss the diminished probability of parental choice bias in selecting childcare facilities in the context of Japan's public childcare system. The accredited childcare facilities under analysis are public services, with local governments assuming administrative responsibilities. When parents wish to use a childcare facility, they first compile a list of preferred facilities and submit it to the local government, with their preferences ranked in order. The local government allocates spaces by considering the lists parents submit in conjunction with availability. Not all parental preferences are guaranteed as the allocation is determined by an algorithm based on the number of applicants and available slots. There is no parental discretion in the allocation of accessible facilities, and there is an element of randomness in the

allocation process. Thus, parental selection bias is less likely to occur. Further, we provide three pieces of evidence against parental selection bias. First, we report that the results from regressions of predetermined characteristics before the choice of childcare on ECERS-3 are statistically insignificant, except for the significant but negligible association of birth weight (see Appendix Table A2).

Second, we examined the association between parents' ratings of childcare centers and corresponding ECERS-3 scores using data from a review conducted by the Tokyo Metropolitan Foundation of Social Welfare and Public Health; no association was found between the two (Fujisawa et al. 2023b). Moreover, no association was observed between parents' ratings of childcare quality on an internet site (Kindergarten and Nursery School Information for Everyone, 2023) and corresponding ECERS-3 total scores (Coef. = −0.003, $p = 0.476$). Thus, parents are unaware of childcare quality, which is consistent with previous findings (Herbst et al., 2020; Mocan, 2007).

Third, we demonstrated that the within-center variance in ECERS-3 across classrooms (or entry years) was larger than the between-center variance (N = 46) from FY2020 to FY2022 in the target city (overall variance = 0.570, between-center variance = 0.297, within-center variance = 0.487; see Appendix Table A3 for the variance in the subscales). Even if parents successfully lobby the local government to ensure that their child is placed in a high-quality childcare center, it is difficult for parents to choose the appropriate classroom (or an entry year) for their child. Because the centers cannot access information on parental socioeconomic status, they cannot use such information to assign children to classrooms. Additionally, school entry timing in Japan is determined by children's birthdate, and delaying school entry or redshirting is not allowed.

From these findings, we assumed that parental selection based on childcare quality did not drive our empirical results. Thus, the between-center comparisons based on ordinary least squares regressions were suitable to examine the causal relationship between ECEC quality and child and parent outcomes.

Main results

Our main results are presented in Table 2. The magnitude of coefficients across models does not change significantly when we include additional controls (Model 2), indicating that our estimate is less likely to suffer from omitted variable bias. The results did not change when the missing values were imputed (Model 3) or when only children with data at both age 5 and Grade 2 were included (Model 4).

As shown in Table 2, we found no significant causal relationship between childcare quality and children's outcomes at age 5 years measured using KIDS and SDQ. However, childcare quality positively affected parental feelings and mental health.

The association between childcare quality and children's outcomes after school entry was substantial. Although the results of the standardized math test scores were suggestive, a one-point increase in ECERS-3 scores at age 5 years raised standardized reading scores at Grade 2 by 0.377–0.554 SD. In addition, the hours of studying that may proxy students' studying habits or diligence is statistically significant at the 10% level. The Grade 2 behavioral problem coefficient is statistically significant at the 5% level with a Model 1 specification.

Notably, regardless of child and parent outcomes before or after school entry, no heterogeneous effect in terms of gender, birth weight, or parental socioeconomic status was found in a consistent manner (see Appendix Table A4). This may be good for policymakers because there is no concern that childcare quality has disproportionally affected a particular

subgroup of children. In addition, we found no systematic pattern of the effect of subscales on outcomes (see Appendix Table A5). These results suggest that outcomes may be influenced by overall childcare quality rather than any particular dimensions of quality. Thus, ECEC quality in Japan has a short-term effect on parental outcomes and a medium-term effect on children's academic outcomes.

Table 2

Relation between ECERS-3 total score at age 5 and child and parent outcomes

	Child outcomes at age 5								Parental outcomes at age 5			
	KIDS				Behavioral problems				Discipline			
Model	1	2	3	4	1	2	3	4	1	2	3	4
ECERS-3	0.082	−0.869	−0.482	−0.070	−1.751	−0.747	−1.434	−1.199	−1.483	−1.412^{+}	−1.441	−1.653
	(5.475)	(5.888)	(5.636)	(5.672)	(1.133)	(0.909)	(1.043)	(1.172)	(1.025)	(0.735)	(1.003)	(1.005)
CC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
PC		✓	✓			✓	✓			✓	✓	
N	197	157	197	158	192	156	192	153	190	154	190	151
R^2	0.161	0.267		0.164	0.079	0.176		0.061	0.028	0.059		0.044
Adjusted R^2	0.135	0.222		0.130	0.050	0.125		0.022	−0.004	0.000		0.004
	Parent outcomes at age 5 (*cont.*)											
	Positive feeling				Negative feeling				Mental health			
Model	1	2	3	4	1	2	3	4	1	2	3	4
ECERS-3	1.025**	0.972*	0.973**	0.842*	1.724**	1.397*	1.660**	1.654**	2.462**	3.052**	2.314**	2.739**
	(0.260)	(0.373)	(0.251)	(0.290)	(0.489)	(0.486)	(0.514)	(0.490)	(0.646)	(0.704)	(0.648)	(0.888)
CC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PC		✓	✓			✓	✓			✓	✓	
N	195	158	195	155	192	156	192	153	190	156	190	152
R^2	0.043	0.094		0.034	0.029	0.056		0.037	0.053	0.082		0.065
Adjusted R^2	0.012	0.039		−0.005	−0.002	−0.003		−0.002	0.022	0.026		0.026
	Child outcomes at Grade 2											
	Math			Reading			Study hours			Behavioral problems		
Model	1	2	3	1	2	3	1	2	3	1	2	3

ECERS-3	0.364	0.522*	0.293	0.483*	0.554*	0.377*	0.234	0.318+	0.222+	−2.909*	−1.438	−2.499
	(0.210)	(0.202)	(0.164)	(0.174)	(0.159)	(0.127)	(0.123)	(0.146)	(0.102)	(1.125)	(1.259)	(1.301)
CC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PC		✓	✓		✓	✓		✓	✓		✓	✓
N	159	129	159	159	129	159	116	91	116	116	92	116
R^2	0.114	0.229		0.128	0.269		0.116	0.134		0.108	0.106	
Adjusted R^2	0.079	0.171		0.094	0.214		0.067	0.038		0.058	0.008	

Note. +: $p < 0.10$. *: $p < 0.05$. **: $p < 0.01$. Model 1 specifications controlled for children's demographic characteristics, such as gender, birth month, and birth weight. Model 2 additionally controlled for parent characteristics, such as parents' years of education and household income. Model 3 imputes the missing values using the multiple imputation method. Robust standard errors are shown in parentheses. Model 4 included only children with data at both age 5 years and Grade 2. The number of observations varied by models because not all parents necessarily answered all questions in the parent survey and not all children who were enrolled in childcare centers at age 5 enrolled in elementary schools in the same city. ECERS-3: ECERS-3 total score. CC: child characteristics are controlled. PC: parent characteristics are controlled.

Study 2

The finding that high-quality childcare produces better outcomes leads to the question of how to improve childcare quality. To answer this question, we conducted a clustered randomized controlled trial (cluster-RCT) to evaluate whether the PD program results in a high ECERS-3 score.

Method

Random assignment

Due to the small sample size, sampling might be biased toward centers with low or high capacity despite random selection of centers. The number of children enrolled in a center varied from 41 to 285 (mean 127.823, SD = 73.313). Thus, sampling was stratified by the number of children enrolled in a center (≥100 or <100 children) and the type of entity (kindergarten or accredited childcare centers). The treatment centers were randomly selected from two from three kindergartens, four from eight preschools with 100 or more children, and three from six preschools with less than 100 children. Four centers included two 5-year-old classes, and 13 centers included only one 5-year-old class. Centers with two classes were assigned to either the treatment or control group. As such, eleven and nine classes were assigned to the treatment and control groups, respectively. The directors and caregivers who were in charge of the targeted classes in the treatment group centers were targeted for the PD program. The balance test showed no difference between the treatment and control groups, except that the years of working experience as a caregiver were longer for the control group and the caregivers' education level was higher for the targeted classrooms (see Appendix Table A6).

Intervention procedure

All participating classrooms were evaluated twice for ECEC quality using ECERS-3, once as a baseline assessment from April to June 2022 and once as an endline assessment from January to early March 2023. The mean (SD) number of days between the baseline and endline assessments was 252.50 (21.71) (min = 210, max = 291). Given that there are substantial between- and within-center variations in ECEC quality measured by ECERS-3 (Fujisawa et al., 2023a), center directors and caregivers in charge of the classes should be the targets of the PD program. These center directors and caregivers did not change during the baseline and endline assessments.

An expert and another assessor conducted the classroom baseline assessments of the treatment group. The expert is a nationally licensed childcare teacher and a certified psychologist having several years of experience using ECERS-3 for evaluation. Baseline assessments of the control group and endline assessments of all classrooms were conducted by assessors other than the expert, who were not informed which group the participating classrooms were assigned to.

The expert provided detailed feedback on the baseline assessment results to the center directors and caregivers in charge of the classes assigned to the treatment group. The mean (SD) number of days between the baseline assessment and feedback for each center was 57.91 (23.36) (min = 22, max = 101). Feedback was provided as follows online and required approximately 1 hour for completion.

(1) Introducing the topic: This entailed the explanation of ECERS-3, including its development in the U.S.; academic theories that ECERS-3 is based on; and key evaluation points. The regional comparison of ECERS-3 results conducted in other municipalities in Japan was also presented.

(2) Communicating positive aspects: Positive aspects of caregiver interactions and environmental settings observed during the assessment—not necessarily related to ECERS-3 evaluation criteria—were communicated. Specific statements and concrete examples were provided when possible. This purpose was to establish a rapport between the expert providing feedback, the center director, and the caregiver in charge of the class.

(3) Communicating areas of improvement: This is the key element of the PD program. Indicators for items with a score less than 3 (minimal) were targeted. Feedback was provided only for negatively evaluated indicators among level 1 and 3 indicators for items with a score less than 3. A detailed explanation was provided for the negative evaluation. Descriptions were supplemented with photos or other materials to explain environmental settings and caregiver interactions. No feedback was provided for items that scored three or higher.

(4) Discussing results and reviewing questions.

(5) Confirming the schedule for endline assessment.

The control group classrooms received no feedback after the baseline assessment. After the RCT, baseline and endline result assessments were sent to the control group.

Statistical analysis

In the subsequent analysis, we examine the changes in ECERS-3 scores induced by feedback interventions through the following regression analysis:

$$ECERS_{ct} = \delta + \tau D_{ct} \times T + \phi_c + \mu_t + u_{ct}.$$

The parameter of interest, denoted as τ, captures the alterations in ECERS-3 scores resulting from the feedback intervention. ECERS-3 assessments were conducted before and after the intervention. Therefore, we can control for the baseline ECERS-3 scores, represented by ϕ_c, enabling us to assess changes based on these scores. This control is crucial, particularly in cases

where the sample size is limited and the baseline scores lack balance. The parameter μ_t encapsulates the average annual changes in scores over one year in situations where no intervention was administered.

Furthermore, to scrutinize the mechanism underlying the effects of the intervention, we conducted analyses based on more refined outcomes and the treatment variable. First, estimations were made for each ECERS-3 subscale, which allowed us to validate the efficacy of our feedback concerning various domains of ECEC quality. Second, the ECERS-3 indicators were systematically categorized, with allowance for overlaps, into domains—including "1. Caregiver's attitude toward children," "2. Object-related," "3. Time use," and "4. Other"—and the percentages of positively assessed indicators in that category were used as outcome variables (see Appendix Table A7). In this estimation, the number of indicators that received a given intervention was used as the treatment variable rather than a binary variable measuring whether intervention was received. This ensured that the number of indicators receiving the intervention differed for each treatment class, which was reflected in the analysis. Finally, via an exploratory analysis aimed at determining the genuine source of our results, we differentiated estimations for indicators that received feedback during the intervention and those that did not.

Results and Discussion

Table 3 shows the descriptive statistics for control and treatment groups' pre- and postintervention ECERS-3 scores. In the treatment group, the mean score, with a preintervention score below 4, surpassed score 4 postintervention. This growth in ECERS-3 scores, absent in the control group, indicates enhanced childcare quality due to our feedback. There was a discrepancy between the preintervention scores of the control and treatment groups, with the latter's scores being marginally lower. This difference may be due to the random selection of feedback

recipients, a discrepancy likely arising from the small sample size. Subsequent analyses scrutinized the feedback effects, particularly the ECERS score changes, considering the preintervention baseline.

Table 3

Descriptive statistics for baseline and endline ECERS-3 scores for control and treatment groups

	Treatment group (N = 11)		Control group (N = 9)	
Variable	Base line Mean (SD)	End line Mean (SD)	Base line Mean (SD)	End line Mean (SD)
Total score	3.799(0.609)	4.179(0.335)	4.029(0.439)	3.882(0.432)
Space and Furnishings	3.844(0.853)	4.156(0.488)	4.19(0.707)	4.048(0.655)
Personal Care Routines	4.455(0.688)	5.091(0.875)	4.778(0.341)	4.861(1.032)
Language and Literacy Learning	3.545(0.926)	3.691(0.797)	3.578(0.79)	3.378(0.94)
Activities	2.491(0.468)	2.755(0.476)	2.767(0.464)	2.289(0.607)
Interaction	5.686(0.985)	6.109(0.628)	5.644(1.17)	5.578(0.961)
Program Structure	4.636(1.206)	5.364(1.048)	4.926(1.321)	5.519(0.747)

Table 4 summarizes the estimated results of the equation. The first column of Table 4 indicates that the ECERS-3 total score increased by 0.527 due to the feedback, a statistically significant result. Considering the SD of approximately 0.6 for the preintervention baseline scores, it is evident that the feedback intervention resulted in a 0.88 SD improvement in ECERS-3 total scores. Analyses were performed for ECERS-3 subscales, with statistically significant "Activities" subscale results, as seen in Table 4. As other subscales had fewer treated indicators, their improvements may not be meaningful.

Next, we categorized the indicators targeted by feedback into "Caregiver's attitude toward children," "Object-related," "Time use," and "Other," allowing overlap, and examined the intervention effects for each category and assessed whether improvement was observed. In this

analysis, outcomes were derived from binary variables for each indicator, averaged across categories. The treatment variable refers to the number of indicators given in feedback, which differs across facilities, to capture treatment intensity. Columns (1)–(4) of Table 5 present discernible improvement, particularly in the indicators of "Caregiver's attitude toward children," "Object-related," and "Time use," which were substantiated at the conventional level. Similar results were obtained upon analyzing binary variables regarding whether the patient received intervention as the treatment variable (see Appendix Table A8).

Based on the results, the estimated effect on "Attitude" is 0.049. There are 16 indicators in this category, and the improvement rate per feedback indicator is calculated as (0.049)/(1/16) = 0.784. Thus, when indicators in this category receive feedback, we obtain a compliance rate of 78.4% on average. In a similar way, the compliance rate for "Object related" is 54% and that for "Time use" is 57.6%. For the categories in which improvement is observed, more than half of the indicators in those categories received feedback demonstrating improvement. Particularly noteworthy is the higher rate of improvement in "Attitude" compared with "Object-related" and "Time use," and this difference was statistically significant.

Table 4

Effects of the professional development program on the ECERS-3 total and subscale scores.

	Total score	Space and Furnishings	Personal Care Routines	Language and Literacy Learning	Activities	Interaction	Program Structure
Number of treated indicators	63	14	4	8	29	1	7
Treat × Endline	0.527**	0.455	0.553	0.345	0.741**	0.489	0.135
	(0.183)	(0.330)	(0.404)	(0.344)	(0.229)	(0.310)	(0.530)
Endline	−0.147	−0.143	0.0833	−0.200	−0.478*	−0.0667	0.593
	(0.136)	(0.245)	(0.300)	(0.255)	(0.170)	(0.230)	(0.393)
Constant	3.192**	3.916**	4.182**	2.227**	2.118**	4.089**	3.970**
	(0.213)	(0.384)	(0.469)	(0.400)	(0.266)	(0.360)	(0.616)
Mean	3.974	4.054	4.794	3.555	2.58	5.769	5.1
Observations	40	40	40	40	40	40	40
R-squared	0.830	0.728	0.698	0.808	0.780	0.873	0.744
Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. *: $p < 0.05$. **: $p < 0.01$. Standard errors are shown in parentheses.

Table 5

Treatment effects for different domains.

	(1) Attitude	(2) Object	(3) Time	(4) Other
Total number of treated indicators	16	30	18	9
Number of treated indicators × Endline	0.049***	0.018***	0.032***	0.025
	(0.013)	(0.004)	(0.007)	(0.019)
Endline	−0.095**	−0.092**	−0.044	−0.034
	(0.039)	(0.032)	(0.037)	(0.032)
Constant	0.573***	0.517***	0.559***	0.745***
	(0.074)	(0.056)	(0.068)	(0.067)
Observations	40	40	40	40
R-squared	0.835	0.747	0.803	0.832
Controls	Yes	Yes	Yes	Yes

Note. Attitude: Caregiver's attitude toward children. Object: Object-related. Time: Time-use. **: $p < 0.01$. ***: $p < 0.001$. Standard errors are shown in parentheses.

We also investigated whether the quality of childcare improved as a result of the intervention. First, we divided the indicators in the treatment group into two categories: the treated category, which received feedback, and the non-treated category, which did not. We then evaluated the intervention's effectiveness by comparing the proportion of indicators suggesting improved quality between the treated and non-treated categories in both the treatment and control groups, before and after the intervention. In this analysis, outcomes were derived from binary variables for each evaluation indicator. Column (5) of Appendix Table A8 indicates that an average of 11.3% of the feedback-received indicators exhibited statistically significant improvement. Conversely, column (6), which represents evaluation indicators without feedback, indicates a lack of significant improvement in scores between the control and treatment groups.

Therefore, our intervention's influence in improving quality may be attributable to the evaluation indicators that underwent feedback. The extensive analyses showed that feedback intervention resulted in a significant improvement in quality. Regarding the robustness of our results, we assert that our results are not mere consequences of regression toward the mean. Within the framework of our feedback design, we advised and intervened in evaluation indicators that fell below the minimal standards, i.e., subpar quality. If regression toward the mean were in effect, an improvement in these indicators would be observed. To investigate this, we conducted a placebo experiment that selected indicators that should have undergone intervention within the control group under our feedback design, excluding the actually intervened indicators within the treatment group, and utilized these evaluation indicators as outcomes. The estimated value shows a closely approximated 0.04, which lacks statistical significance (see Appendix Table A9). Thus, the improvement in quality due to feedback cannot be explained by regression toward the mean.

Further, the accuracy of our study in estimating standard errors is somewhat constrained due to the small sample size. Despite the lack of asymptotic precision, we conducted verifications using clustering of errors within facilities (see Appendix Table A10). While standard errors are larger due to the serial correlation of error terms, the discussions about statistical significance have not drastically changed, implying the reliability and robustness of our findings.

Summary and Concluding Discussion

This study examined whether ECEC quality affects child development and parental status using data on the ECEC quality of all childcare centers in one municipality in Tokyo. It also assessed whether a concise PD program can improve ECEC quality. This study considered internal and external validity of the effects of ECEC quality. Previous studies that showed the

effects of ECEC quality had unique samples, such as children growing up in a disadvantaged environment; thus, the external validity of the effect of ECEC quality has not been warranted (Blau, 2021). In addition, there was insufficient verification of selection bias in the context of nonuniversal access to ECEC, which affected the internal validity of the effect of ECEC. The present study investigated ECEC quality where the demand for childcare was met by the childcare facilities, indicating that universal access to ECEC had been almost achieved. Furthermore, selection bias was found to be minimal. Thus, it can be argued more strongly that the association between ECEC quality and outcomes was causal.

To our knowledge, this study is the first to show the effects of childcare quality in Japan's universal ECEC context. In addition, our study complements the results obtained in other countries, such as the U.S. (Peisner-Feinberg, et al. 2001) and Peru (Araujo, et al. 2019). Although the discussion of statistical significance should be considered carefully because of the small sample, ECEC quality in the 5-year-old class had no apparent impact on child outcomes at age 5. However, its impact on parents was positive, such as parents' positive feelings toward their children and improved mental health. While there was no apparent impact on problem behaviors, improved ECEC quality resulted in higher academic performance in Grade 2. The magnitude of our estimates on academic performance appears to be large compared to previous studies; Kraft (2020) reported a median impact of 0.10 SDs across 747 educational interventions to improve children's cognitive skills.

It is worth discussing why the effect of ECEC quality was not apparent in developmental outcomes simultaneously but the substantial effect of ECEC quality was observed in academic performance in the medium term. This result is inconsistent with previous meta-analyses showing a small, albeit positive, association between ECEC quality measured using observation-

based instruments (Brunsek et al., 2017; Perlman et al., 2016) and the recent trend toward more minor effects of childcare on academic outcomes in the U.S. (Whitaker et al.,2023).

This inconsistency may be related to the child-centered focus of Japanese childcare providers. Although not the main focus of the present study, the results of a survey of child caregivers' beliefs conducted prior to the ECERS-3 observation showed that child caregivers in the city focused more on children than on basic skills. Moreover, this child-centered orientation was negatively correlated with the basic-skills orientation (see Appendix A11). This seems to contradict the situation in the U.S.; the increased emphasis on school-like educational content related to basic skills in a teacher-directed manner in recent preschool programs in the U.S. may have reduced the effect of preschool programs on academic outcomes (Whitaker et al.,2023). The child-centered beliefs of Japanese caregivers may not seem straightforward for enabling children to acquire academic skills. However, it may have led to high academic achievement through some process related to the ECEC quality in Japan.

There are some possible reasons why ECEC quality did not simultaneously affect child outcomes. First, the children in the present sample had average KIDS scores above 100 and had fewer problem behaviors than those of other large samples in Japan. Furthermore, ECEC quality in the city was higher than that of other municipal facilities in Japan. Thus, it is possible that we could not successfully identify the association between child outcomes and ECEC quality. Second, child outcomes measured using KIDS and SDQ did not successfully capture child outcomes in the preschool period induced by high-quality ECEC.

The positive impact of ECEC quality on parents may be related to the unique characteristics of Japanese ECEC context as emphasizing support for parents. High-quality ECEC may lead to good relationships between parents and caregivers, thereby improving

parent's mental health. This should be investigated in future research as previous studies' findings were inconsistent; some studies have found that high childcare quality is related to good communication between parents and caregivers (Owen et al., 2000), whereas others have found little connection (Perlman & Fletcher, 2012). Another possibility is that high-quality ECEC affects child development, unmeasured in this study, thus improving parental mental health.

Whether a mediating factor affects the influence of ECEC quality on parents or whether the quality directly affects parents remains unclear. However, parental mental health affects child development (Tung et al., 2023). Therefore, the causal relationship between ECEC quality enhanced academic achievement after school entry may have been mediated by the positive outcome of parents. For example, Kim et al. (2018) state that parents' positive engagement with a child during preschool years had a causal impact on adolescent outcomes 10 years later.

Only if ECEC quality measured by ECERS-3 is shown to have causal effects on outcomes does it make sense to target the ECERS-3 measure to improve ECEC quality. In Study 1, we showed that ECEC quality as measured by ECERS-3 had a causal effect, and in Study 2, we tested the effect of PD that targeted the evaluation indicators of ECERS-3.

The PD program developed in this study was found to significantly improve ECEC quality. Indicators with detailed feedback were significantly improved compared to those without. The effect of the PD program of 0.88 SD in the ECERS-3 total score is larger than the meta-analysis report, showing the pooled effect size of the effectiveness of the PD program that included the coaching component for teachers was about 0.5 SD (Kraft, Blazar, & Hogan, 2018).

By observing the degree of improvement across subscales, we expected to identify ECEC quality domains that were most affected by the intervention. The "Activities" subscale showed significant improvement. However, as several evaluation indicators that received intervention

were included in the this subscale and the evaluation indicators in other subscales were not subject to much intervention, examining the degree of improvement by intervention for each subscale did not produce meaningful results. Instead, we categorized the lowest level of ECERS-3, the indicator level, in relation to its content to determine the intervention effects across the domains of "Caregiver's attitude," "Object-related," and "Time use."

The contents related to material aspects, such as providing appropriately sized furniture and making educational materials or toys accessible to children, were expected to improve relatively easily because they do not directly get into the inner beliefs of caregivers. Further, the Japanese government subsidizes childcare facilities to purchase necessary items based on the age and number of children. As expected, significant improvement was confirmed for the object-related domain. In contrast, caregiver's attitudes, such as those reflected in interactions between the caregiver and the child, were expected to be difficult to improve as such a modification would require changes in the caregivers' beliefs and thoughts. However, the caregiver's attitude domain was also improved by the intervention. These results suggest that the PD program in this study is effective in areas that might be challenging to change through intervention.

The PD program implemented in this study was concise, with the expert providing feedback on the ECERS-3 results in approximately 1 hour. However, because the feedback was focused on items with less than 3, it was individualized and tailor-made for caregivers who needed to know what needed to be improved. Compared to programs in which the center director and consultant jointly choose the intervention based on self-assessment (e.g., Helmerhorst, et al., 2017), the target of the PD program in this study was determined systematically, so it is scalable whether the target has improved. Therefore, the program was effective despite being shorter time than in previous studies. Furthermore, the validation results of the time-saving PD program in

Japanese ECEC make a significant contribution to the insights gained from PD program validation studies, which have been conducted predominantly in Europe and the U.S. (e.g., Egert et al., 2018:2020; Ragni et al., 2021).

Limitations

This study has certain limitations. First, the effect of measurement error in ECERS-3 cannot be ruled out. Second, the ECEC quality of the childcare centers was generally high compared to a previous report conducted in a different city in Japan (Fujisawa et al., 2023a). Third, the sample size was small, although the present study targeted all childcare centers in the city and children. These limitations should be considered when interpreting the results of Study 1 and Study 2.

Any estimates in Study 1 might have been affected by the measurement error in ECERS-3 scores, the small sample size, and the relatively high scores of ECERS-3 in the city. Due to these limitations, the variations in variables may be too small to detect individual differences in the relationship between childcare quality and child development. Another limitation in Study 1 was the cohort of a 5-year-old class of children who experienced the COVID-19 pandemic in 2020. Thus, results may be related to the lack of a clear relationship between ECEC quality and child outcomes at age 5 in the present study. The results of a different longitudinal study in Japan reported that children who experienced the COVID-19 pandemic showed developmental delays of several months at age 5 and that individual development differences widened throughout the pandemic (Sato et al., 2023). Another limitation of Study 1 was that the selection of unobservable variables was not verified.

Study 2 also has limitations. The small sample size and the study design did not allow us to examine the extent to which the variation resulting from the intervention was due to changes

in the caregivers and/or to changes at the center level. Moreover, examining the spillover effect within the facilities was impossible, both among classes that received the PD program and those that were not targeted through communication with the center directors and caregivers. Future work will need to address these issues. The high ECERS-3 score in the facilities in this city could also be explained by the skills of caregivers and their ability to understand feedback and apply it to their practice. It is necessary to verify whether the PD program developed in this study positively affects classes with very low quality.

Despite these limitations, this study demonstrated a causal relationship between ECEC quality and child and parent outcomes and conducted an RCT to validate the PD program for improving childcare quality. This study contributed to the literature on childcare quality, which had been regionally limited, and provided evidence for the need to improve childcare quality through policy and financial investments.

References

Araujo, M. C., Dormal, M., & Schady, N. (2019). Childcare Quality and Child Development. *Journal of Human Resources*, *54*(3), 656–682. https://doi.org/10.3368/jhr.54.3.0217.8572R1

Blau, D. M. (2021). The Effects of Universal Preschool on Child and Adult Outcomes: A Review of Recent Evidence from Europe with Implications for the United States. *Early Childhood Research Quarterly*, *55*, 52–63. https://doi.org/10.1016/j.ecresq.2020.10.009

Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLOS ONE*, *12*(6), e0178512. https://doi.org/10.1371/journal.pone.0178512

Brunsek, A., Perlman, M., McMullen, E., Falenchuk, O., Fletcher, B., Nocita, G., Kamkar, N., & Shah, P. S. (2020). A meta-analysis and systematic review of the associations between professional development of early childhood educators and children's outcomes. *Early Childhood Research Quarterly*, *53*, 217–248. https://doi.org/10.1016/j.ecresq.2020.03.003

Burchinal, M. (2018). Measuring Early Care and Education Quality. *Child Development Perspectives*, *12*(1), 3–9. https://doi.org/10.1111/cdep.12260

Burchinal, M., Whitaker, A. A., & Jenkins, J. M. (2022). The promise and purpose of early care and education. *Child Development Perspectives*, *16*(3), 134–140. https://doi.org/10.1111/cdep.12463

Children and Family Agency, Japan (2023). Summary of the situation related to childcare centers. https://www.cfa.go.jp/policies/hoiku/torimatome/r5/ [In Japanese]

Deater-Deckard, K. (2000). Parenting and Child Behavioral Adjustment in Early Childhood: A Quantitative Genetic Approach to Studying Family Processes. *Child Development*, *71*(2), 468–484. https://doi.org/10.1111/1467-8624.00158

Egert, F., Dederer, V., & Fukkink, R. G. (2020). The impact of in-service professional development on the quality of teacher-child interactions in early education and care: A meta-analysis. *Educational Research Review*, *29*, 100309. https://doi.org/10.1016/j.edurev.2019.100309

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, *88*(3), 401–433. https://doi.org/10.3102/0034654317751918

Fort, M., Ichino, A., & Zanella, G. (2020). Cognitive and Noncognitive Costs of Day Care at Age 0–2 for Children in Advantaged Families. *Journal of Political Economy*, *128*(1), 158–205. https://doi.org/10.1086/704075

Fujisawa, K. K., Fukai, T., & Nakamuro, M. (2023a). Quality of early childhood education and care in Japanese accredited nursery centers: A study using the Early Childhood Environment Rating Scale, Third Edition (ECERS-3). *PLOS ONE*, *18*(2), e0281635. https://doi.org/10.1371/journal.pone.0281635

Fujisawa, K. K., Sugita, S., Fukai, T., & Nakamuro, M.(2023b). The relationship between third-party evaluation of welfare services and the quality of childcare: current status and issues. RIETI Discussion Paper Series 22-J-042 [In Japanese]

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2021). Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly*, *54*, 1–12. https://doi.org/10.1016/j.ecresq.2020.07.009

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised.* New York: Teachers College Press.

Harms, T., Clifford, R., & Cryer, D. (2015). *Early Childhood Environment Rating Scale (3rd ed).* New York, NY: Teachers College Press

Hashimoto, K. (2013). Validity of the Family-Rated Kinder Infant Development Scale (KIDS) for Children. *Pediatrics & Therapeutics*, *03*(02). https://doi.org/10.4172/2161-0665.1000153

Havnes, T., & Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, *127*, 100–114. https://doi.org/10.1016/j.jpubeco.2014.04.007

Helmerhorst, K. O. W., Fukkink, R. G., Riksen-Walraven, J. M. A., Deynoot-Schaub, M. J. J. M. & Tavecchio, L W. C. (2017). Improving quality of the child care environment through a consultancy programme for centre directors, *International Journal of Early Years Education*, *25*(4), 361-378. doi: 10.1080/09669760.2017.1321528

Herbst, C. M., Desouza, K. C., Al-Ashri, S., Srivatsav Kandala, S., Khullar, M., & Bajaj, V. (2020). What do parents value in a child care provider? Evidence from Yelp consumer

reviews. *Early Childhood Research Quarterly*, *51*, 288–306. https://doi.org/10.1016/j.ecresq.2019.12.008

Iida, Y., Moriwaki, A., Komatsu, S., & Kamio, Y. (2014). Standardization of the Strengths and Difficulties Questionnaire among Japanese 4-5 year-old preschool children, In: Kamio, Y., ed. *Annual report of research supported by health and labour sciences research grants.* National Center of Neurology and Psychiatry, Tokyo, pp.33-41. [In Japanese].

Japel, C., Tremblay, R. E., & Côté, S. (2005). *Quality Counts! Assessing the Quality of Daycare Services*. IRPP. https://irpp.org/research-studies/quality-counts/

Kindergarten and Nursery School Information for Everyone (2023). *Review Ranking Childcare Centers Section.* https://www.minkou.jp/kinder/ranking/review/ [In Japanese]

Kim, J. H., Schulz, W., Zimmermann, T., & Hahlweg, K. (2018). Parent–child interactions and child outcomes: Evidence from randomized intervention. *Labour Economics*, *54*, 152–171. https://doi.org/10.1016/j.labeco.2018.08.003

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of educational research*, *88*(4), 547-588. https://doi.org/10.3102/0034654318759268

Markowitz, A. J., & Ansari, A. (2020). Changes in academic instructional experiences in Head Start classrooms from 2001–2015. *Early Childhood Research Quarterly*, *53*, 534-550. https://doi.org/10.1016/j.ecresq.2020.06.008

Ministry of Education, Culture, Sports, Science and Technology, Japan (2020). *Survey of the actual conditions of preschool education*. [in Japanese].

Ministry of Health, Labour and Welfare, Japan. (2018). Nursery Center Childcare Guidelines [in Japanese]. Ministry of Health, Labour and Welfare, Japan.

Mocan, N. (2007). Can consumers detect lemons? An empirical analysis of information asymmetry in the market for child care. *Journal of Population Economics*, *20*(4), 743–780. https://doi.org/10.1007/s00148-006-0087-6

Moriwaki A., & Kamio, Y. (2014). Normative data and psychometric properties of the Strengths and Difficulties Questionnaire among Japanese school-aged children. *Child and Adolescent Psychiatry and Mental Health*, *8*(1). https://doi.org/10.1186/1753-2000-8-1.

Ochanomizu University Center for Women's Education and Development. (2006). *The history of Japan's preschool education and care.* https://www.ocha.ac.jp/intl/cwed_old/eccd/report/history2006_E.pdf

OECD (2019). *TALIS Starting Strong 2018 Technical Report.* OECD Publishing, Paris. https://www.oecd.org/education/talis/TALIS-Starting-Strong-2018-Technical-Report.pdf

OECD. (2023). Enrolment in childcare and pre-school. In OECD FAMILY DATABASE. https://www.oecd.org/els/family/database.htm

Oppermann, E., Lehrl, S., & Burghardt, L. (2023). Associations between preschool quality and children's social-emotional development until 2nd grade of elementary school. *Early Childhood Research Quarterly*, *63*, 133–144. https://doi.org/10.1016/j.ecresq.2022.12.002

Owen, M. T., Ware, A. M., & Barfoot, B. (2000). Caregiver-mother partnership behavior and the quality of caregiver-child and mother-child interactions. *Early Childhood Research Quarterly*, *15*(3), 413–428. https://doi.org/10.1016/S0885-2006(00)00073-9

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & 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. https://doi.org/10.1111/1467-8624.00364

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A Systematic Review and Meta-Analysis of a Measure of Staff/Child Interaction Quality (the Classroom Assessment Scoring System) in Early Childhood Education and Care Settings and Child Outcomes. *PLOS ONE*, *11*(12), e0167660. https://doi.org/10.1371/journal.pone.0167660

Perlman, M., & Fletcher, B. A. (2012). Hellos and How Are Yous: Predictors and Correlates of Communication Between Staff and Families During Morning Drop-Off in Child Care Centers. *Early Education and Development*, *23*(4), 539–557. https://doi.org/10.1080/10409289.2010.548766

Ragni, B., Boldrini, F., Buonomo, I., Benevene, P., Grimaldi Capitello, T., Berenguer, C., & De Stasio, S. (2021). Intervention Programs to Promote the Quality of Caregiver–Child Interactions in Childcare: A Systematic Literature Review. *International Journal of Environmental Research and Public Health*, *18*(21), Article 21. https://doi.org/10.3390/ijerph182111208

Sato, K., Fukai, T., Fujisawa, K. K., & Nakamuro, M. (2023). Association Between the COVID-19 Pandemic and Early Childhood Development. *JAMA Pediatrics*, *177*(9), 930–938. https://doi.org/10.1001/jamapediatrics.2023.2096

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, *9*(2), 109–124. https://doi.org/10.1177/1476718X10387900

Tung, I., Hipwell, A. E., Grosse, P., Battaglia, L., Cannova, E., English, G., Quick, A. D., Llamas, B., Taylor, M., & Foust, J. E. (2023). Prenatal stress and externalizing behaviors in childhood and adolescence: A systematic review and meta-analysis. *Psychological Bulletin*, 10.1037/bul0000407. https://doi.org/10.1037/bul0000407

Tokyo Metropolitan Government (2022). *Results of the 2022 Survey of Childcare Workers in Tokyo.* https://www.fukushi.metro.tokyo.lg.jp/kodomo/shikaku/r4hoikushichousa.html [In Japanese]

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network. (2010). Do Effects of Early Child Care Extend to Age 15 Years? Results From the NICHD Study of Early Child Care and Youth Development. *Child Development*, *81*(3), 737–756. https://doi.org/10.1111/j.1467-8624.2010.01431.x

Yamaguchi, S., Asai, Y., & Kambayashi, R. (2018). How does early childcare enrollment affect children, parents, and their interactions? *Labour Economics*, *55*, 56–71. https://doi.org/10.1016/j.labeco.2018.08.006

Whitaker, A. A., Burchinal, M., Jenkins, J. M., Bailey, D. H., Watts, T. W., Duncan, G. J., Hart, E. R., & Peisner-Feinberg, E. S. (2023). Why are Preschool Programs Becoming Less Effective? (EdWorkingPaper: 23-885). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/smqa-n695

Acknowledgement

The authors would like to extend their sincerest appreciation to the mayor of the municipality, the staff members of the respective department, the directors and caregivers in the participating childcare centers and kindergartens, and the assessors for their cooperation with the data collection for this study. A special thank you to Professor Reiko Uzuhashi for the immense advice and guidance during the ECERS-3 data collection process and Ms. Ayu Akitomi for managing our research project.

This study is conducted as a part of the Project "Implementing Evidence-Based Policy Making in Japan" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The following research grants also supported this paper: the Tokyo Foundation for Policy Research, the Research Institute of Science and Technology for Society (RISTEX; JPMJRX21B1), Grants-in-Aid for Scientific Research (KAKENHI; 18H00997, 22H00064, 23K12487), Keio University Academic Development Funds, Keio Gijuku Fukuzawa Memorial Fund for the Advancement of Education and Research, the Grants for Social Welfare Activities, the Mitsubishi Foundation (ID202130031).

This study was approved by the Keio University SFC Ethics Committee for Experiments and Research (Reception No. 290).

Appendix

Table A1

Comparison of descriptive statistics for the present study and the Fujisawa et al. (2023) of the ECERS-3 subscales.

	Present study			Fujisawa et al (2023)		
ECERS-3	Mean	SD	N	Mean	SD	N
Total	4.605	0.351	16	3.41	0.69	88
Space and furnishings	4.571	0.833	16	3.50	0.86	88
Personal care routines	5.600	0.944	16	4.30	1.21	88
Language and literacy	4.170	0.948	16	3.36	0.92	88
Learning activities	3.187	0.484	16	2.21	0.68	88
Interaction	6.500	0.502	16	4.76	1.24	88
Program structure	5.778	0.851	16	4.01	1.44	88

Note. The unit of observations is classroom. The score in the present study is calculated by using the data from the fiscal year of 2020 in the target city. The score of Fujisawa, et al. (2023) is calculated by using representative data from the fiscal year of 2017 to 2019 in another local city of the Tokyo metropolitan area.

Table A2

The relationship between predetermined characteristics and ECERS-3.

	ECERS-3 Total	Space and furnishings	Personal care routines	Language and literacy	Learning activities	Interaction	Program structure
Father's education (in years)	-0.004	0.006	0.011	0.020	-0.022	-0.016	0.001
	(0.013)	(0.048)	(0.039)	(0.033)	(0.018)	(0.014)	(0.034)
Mother's education (in years)	-0.009	0.009	-0.053+	0.015	-0.012	-0.022	-0.009
	(0.010)	(0.017)	(0.029)	(0.029)	(0.017)	(0.019)	(0.025)
Income	0.000	0.000	0.000	0.000	0.000^{+}	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender (female=1)	-0.033	-0.058	0.011	-0.243	0.026	-0.043	0.116
	(0.052)	(0.076)	(0.121)	(0.139)	(0.091)	(0.071)	(0.096)
Birth weight (>2,500g)	-0.178+	0.374^{+}	-0.524+	-0.432	-0.190	-0.290	-0.556*
	(0.085)	(0.212)	(0.262)	(0.351)	(0.172)	(0.172)	(0.229)
Birth of month (July-Sep=1)	-0.025	0.050	-0.080	-0.123	-0.002	-0.097	0.076
	(0.052)	(0.129)	(0.135)	(0.205)	(0.085)	(0.068)	(0.150)
Birth of month (Oct-Dec=1)	0.027	0.055	-0.110	-0.170	0.088	0.077	0.153
	(0.063)	(0.150)	(0.200)	(0.278)	(0.099)	(0.135)	(0.186)
Birth of month (Jan-Mar=1)	0.083	0.047	0.093	0.036	0.067	0.091	0.259*
	(0.059)	(0.141)	(0.154)	(0.173)	(0.061)	(0.114)	(0.093)
N	160	160	160	160	160	160	160
R-squared	0.069	0.054	0.061	0.048	0.044	0.065	0.082
Adjusted R-squared	0.020	0.004	0.012	-0.002	-0.006	0.016	0.034

Note. Figures in parentheses are cluster robust standard errors. The reference category for the variable of children's birth of months is ones who were born during April to June. $^{+}$: $p < 0.10$; *: $p < 0.05$.

Table A3

Between-centers and within-centers variances in ECERS-3.

	ECERS-3 Total	Space and furnishings	Personal care routines	Language and literacy	Learning activities	Interaction	Program structure
Overall	0.570	0.820	0.836	0.985	0.623	0.882	1.124
Between	0.297	0.590	0.369	0.720	0.357	0.513	0.629
Within	0.487	0.598	0.750	0.700	0.515	0.718	0.941

Note. The variance is calculated by using the data (N=46) from the fiscal year of 2020 to 2022 in the target city in Tokyo.

Table A4

Heterogenous effects on outcomes in terms of gender, birth weight, and parental socio-economic status.

	At Age 5						At Grade 2			
	KIDS	Behavioral problems	Discipline	Positive feeling	Negative feeling	Mental health	Math	Reading	Study Hours	Behavioral problems
Gender (female = 1) × total	-5.966	-2.162	-2.881	1.243	2.107	2.453^{+}	-0.895	-0.025	0.238	-1.219
	(8.045)	(2.658)	(2.176)	(0.728)	(2.075)	(1.390)	(0.489)	(0.453)	(0.318)	(3.015)
Total	2.150	0.324	0.051	0.348	0.372	1.822	0.962**	0.567*	0.205*	-0.888
	(3.706)	(1.554)	(0.947)	(0.389)	(0.926)	(1.279)	(0.183)	(0.179)	(0.067)	(1.651)
Gender (female = 1)	40.717	7.427	12.799	-5.546	-9.275	-10.940	4.522+	0.621	-0.883	3.604
	(37.530)	(11.985)	(9.662)	(3.409)	(9.180)	(6.262)	(2.052)	(2.106)	(1.460)	(13.558)
Observations	157	156	154	158	156	156	129	129	91	92
R-squared	0.270	0.180	0.078	0.105	0.063	0.087	0.246	0.269	0.140	0.108
Adjusted R-squared	0.220	0.123	0.014	0.044	-0.001	0.024	0.182	0.207	0.032	-0.002
Low BW (yes = 1) × total	-11.959	2.502	1.657	-3.317*	-1.768	-6.510	0.538	0.021	-0.819	7.609^{+}
	(14.452)	(6.915)	(5.331)	(1.482)	(4.529)	(4.952)	(1.653)	(1.809)	(0.848)	(3.888)
Total	10.495	-3.120	-2.981	4.120*	3.072	9.235+	0.014	0.535	1.090	-8.603+
	(13.759)	(6.668)	(5.537)	(1.464)	(4.563)	(4.364)	(1.708)	(1.839)	(0.874)	(3.840)
Low BW (yes = 1)	58.113	-12.620	-8.853	15.439*	7.875	30.998	-2.122	-0.053	3.985	-36.485
	(70.119)	(32.949)	(24.826)	(6.856)	(21.015)	(23.397)	(8.024)	(8.804)	(4.034)	(18.820)
Observations	157	156	154	158	156	156	129	129	91	92
R-squared	0.270	0.177	0.060	0.111	0.057	0.090	0.230	0.269	0.153	0.127
Adjusted R-squared	0.220	0.120	-0.005	0.051	-0.008	0.027	0.165	0.207	0.047	0.020
Poverty × total	9.983	-2.094	2.965	-2.036	-0.090	-10.356	-0.261	0.782	-0.702	-2.286
	(8.382)	(7.270)	(2.279)	(2.659)	(2.734)	(6.499)	(1.053)	(0.599)	(0.376)	(3.211)

Total	-1.465	-0.632	-1.576*	1.079**	1.402*	3.614**	0.549+	0.516*	0.351*	-1.367
	(5.964)	(1.154)	(0.674)	(0.338)	(0.489)	(0.755)	(0.228)	(0.177)	(0.139)	(1.322)
Poverty (yes = 1)	-46.315	11.724	-14.024	8.964	0.632	48.304	0.327	-3.965	3.248+	11.679
	(37.547)	(34.266)	(10.647)	(11.692)	(12.741)	(29.257)	(4.804)	(2.825)	(1.610)	(14.658)
Observations	157	156	154	158	156	156	129	129	91	92
R-squared	0.269	0.188	0.066	0.101	0.056	0.105	0.267	0.281	0.144	0.112
Adjusted R-squared	0.214	0.126	-0.007	0.034	-0.016	0.037	0.199	0.214	0.025	-0.010

Note. The coefficients are the interaction term added into Model 2 (controlling for children and parent demographic characteristics).

Total: ECERS-3 total score. Low birth weight (BW) is defined as one if his/her birth weight is less than 2500g. Poverty is defined as one if the household annual income is less than 2,500 thousand JPY. Robust standard errors are in parentheses. $^{+}$: $p < 0.10$. *: $p < 0.05$. **: $p < 0.01$.

Table A5

Effects of the subscales of ECERS-3 on outcomes.

	At Age 5			At Grade 2						
	KIDS	Behavioral problems	Discipline	Positive feeling	Negative feeling	Mental health	Math	Reading	Study Hours	Behavioral problems
Space and furnishings	-0.560 (2.600)	-0.212 (0.284)	-0.246 (0.364)	0.045 (0.183)	0.062 (0.291)	0.102 (0.468)	0.035 (0.102)	0.046 (0.090)	-0.054 (0.054)	-0.720 (0.466)
Personal care routines	3.386 (2.628)	-0.207 (0.404)	-0.239 (0.306)	0.043 (0.170)	0.270 (0.207)	0.638* (0.250)	0.001 (0.156)	0.126 (0.120)	0.034 (0.065)	0.420 (0.288)
Language and literacy	0.026 (1.803)	0.106 (0.246)	-0.295 (0.201)	0.280* (0.123)	0.178 (0.192)	0.802** (0.244)	0.193* (0.059)	0.149+ (0.063)	0.102* (0.033)	0.063 (0.442)
Learning activities	-7.277+ (3.807)	-0.456 (0.807)	-0.898* (0.389)	0.778* (0.317)	0.852* (0.372)	1.637* (0.617)	0.186 (0.157)	0.085 (0.140)	0.105 (0.104)	-1.128 (0.862)
Interaction	6.780+ (3.832)	-0.523 (0.400)	0.065 (0.540)	-0.106 (0.270)	0.361 (0.413)	0.075 (0.688)	0.051 (0.169)	0.233 (0.121)	0.280+ (0.123)	-0.597 (0.539)
Program structure	-3.231 (3.170)	0.160 (0.476)	0.022 (0.258)	0.218 (0.233)	0.350+ (0.173)	0.685 (0.487)	0.219 (0.127)	0.140 (0.074)	0.162** (0.039)	-0.095 (0.561)
N	157	156	154	158	156	156	129	129	91	92

Note. The coefficients are the interaction term added into Model 2 (controlling for children and parent demographic characteristics).

Robust standard errors are in parentheses. +: $p < 0.10$. *: $p < 0.05$. **: $p < 0.01$.

Table A6

Descriptive statistics for the treatment and control groups.

	(1) Treatment		(2) Control		(1)-(2) Pairwise t-test
Variable	N	Mean (SE)	N	Mean (SE)	Mean difference
Centers					
# of staff	9	34.111 (2.721)	8	34.500 (2.383)	-0.389
# of children	9	119.778 (19.899)	8	108.875 (16.281)	10.903
# of classes	9	6.444 (0.444)	8	6.000 (0.189)	0.444
Open hours	9	9.667 (0.882)	8	10.312 (0.688)	-0.646
Extra hours beyond 8 hours	9	2.000 (0.577)	8	1.750 (0.491)	0.250
Caregivers in the center					
Female (=1)	65	0.923 (0.033)	43	0.907 (0.045)	0.016
Age	64	44.969 (1.345)	40	48.000 (2.004)	-3.031
License (Yes =1)	65	1.000 (0.000)	43	1.000 (0.000)	.n
Education (B.A. = 1)	65	0.169 (0.047)	43	0.093 (0.045)	0.076
In charge of class (=1)	65	0.292 (0.057)	43	0.326 (0.072)	-0.033
Full-time (=1)	65	0.738 (0.055)	43	0.767 (0.065)	-0.029
Experiences 1 (in years)	63	17.032 (1.338)	41	21.146 (1.909)	-4.115*
Experiences 2 (in years)	64	12.312 (1.187)	39	13.897 (1.675)	-1.585
Caregivers in charge of the class					
Female (=1)	19	0.842 (0.086)	14	1.000 (0.000)	-0.158
Age	19	37.737 (1.672)	12	39.250 (2.236)	-1.513
License (Yes =1)	19	1.000 (0.000)	14	1.000 (0.000)	.n
Education (B.A.=1)	19	0.316 (0.110)	14	0.071 (0.071)	0.244*
Full-time (=1)	19	0.632 (0.114)	14	0.857 (0.097)	-0.226
Experiences 1 (in years)	19	11.474 (1.478)	14	16.071 (2.536)	-4.598
Experiences 2 (in years)	19	7.553 (1.196)	14	10.357 (1.897)	-2.805

Note. *: $p < 0.05$. Working experience 1: working experience as a teacher (in years); Working experience 2: working experience in the current center (in years)

Table A7

Classification of the indicators in ECERS-3 targeted for intervention based on domains.

Subscale	Item	Indicator	Group
1 Space and Furnishings	2 Furnishings for care, play, and learning	1.3	2
1 Space and Furnishings	2 Furnishings for care, play, and learning	3.4	2
1 Space and Furnishings	3 Room arrangement for play and learning	3.1	2
1 Space and Furnishings	3 Room arrangement for play and learning	3.2	2
1 Space and Furnishings	5 Child-related display	1.3	1
1 Space and Furnishings	5 Child-related display	3.2	2
1 Space and Furnishings	5 Child-related display	3.3	1
1 Space and Furnishings	6 Space for gross motor play	1.3	3
1 Space and Furnishings	6 Space for gross motor play	3.1	2,3
1 Space and Furnishings	6 Space for gross motor play	3.2	4
1 Space and Furnishings	7 Gross motor equipment	1.1	2
1 Space and Furnishings	7 Gross motor equipmen	1.3	3
1 Space and Furnishings	7 Gross motor equipment	3.1	3
1 Space and Furnishings	7 Gross motor equipment	3.3	4
2 Personal care routines	9 Toileting/diapering	3.2	2
2 Personal care routines	9 Toileting/diapering	3.3	4
2 Personal care routines	11 Safety practices	1.1	4
2 Personal care routines	11 Safety practices	3.1	4
3 Language and literacy	12 Helping children expand vocabulary	3.3	1
3 Language and literacy	14 Staff use of books with children	1.1	1
3 Language and literacy	14 Staff use of books with children	3.1	1
3 Language and literacy	15 Encouraging children's use of books	3.1	2,3
3 Language and literacy	16 Becoming familiar with print	1.2	1
3 Language and literacy	16 Becoming familiar with print	1.4	1
3 Language and literacy	16 Becoming familiar with print	3.1	2
3 Language and literacy	16 Becoming familiar with print	3.2	1
4 Learning activities	17 Fine motor	3.1	2,3
4 Learning activities	17 Fine motor	3.3	2

4 Learning activities	17 Fine motor	3.4	2
4 Learning activities	18 Art	3.1	2,3
4 Learning activities	19 Music and movement	1.1	2
4 Learning activities	19 Music and movement	3.1	2,3
4 Learning activities	20 Blocks	1.1	2
4 Learning activities	20 Blocks	3.1	2,3
4 Learning activities	20 Blocks	3.2	2
4 Learning activities	20 Blocks	3.3	4
4 Learning activities	21 Dramatic play	1.1	2
4 Learning activities	21 Dramatic play	1.2	2
4 Learning activities	21 Dramatic play	3.1	2,3
4 Learning activities	21 Dramatic play	3.2	1
4 Learning activities	21 Dramatic play	3.3	1
4 Learning activities	22 Nature/science	3.1	2,3
4 Learning activities	22 Nature/science	3.3	2,3
4 Learning activities	23 Math materials and activities	1.2	1
4 Learning activities	23 Math materials and activities	3.1	2,3
4 Learning activities	23 Math materials and activities	3.2	1
4 Learning activities	24 Math in daily events	3.2	1
4 Learning activities	24 Math in daily events	3.3	1
4 Learning activities	25 Understanding written numbers	1.1	2
4 Learning activities	25 Understanding written numbers	3.1	2
4 Learning activities	25 Understanding written numbers	3.2	2
4 Learning activities	25 Understanding written numbers	3.3	1
4 Learning activities	25 Understanding written numbers	3.4	1
4 Learning activities	26 Promoting acceptance of diversity	1.1	2
4 Learning activities	26 Promoting acceptance of diversity	3.1	2
5 Interaction	29 Individualized teaching and learning	3.3	4
6 Program structure	33 Transitions and waiting times	1.4	3
6 Program structure	34 Free play	1.1	3
6 Program structure	34 Free play	1.3	3
6 Program structure	34 Free play	3.1	3

6 Program structure 6	34 Free play	3.3	3
6 Program structure	34 Free play	3.4	4
6 Program structure 6	35 Whole-group activities for play and learning	3.1	4

Note. In "Group" cell, 1 = "Caregiver's attitude toward children", 2 = "Object-related", 3 = "Time-use", and 4 = "Other".

Table A8

Treatment effect for respective domains and treated/nontreated indicators.

	(1) Attitude	(2) Object related	(3) Time use	(4) Other	(5) Treated	(6) Nontreated
Total number of treated indicators	16	30	18	9	63	132
Treat × Endline	0.161^{+} (0.084)	0.144* (0.063)	0.109 (0.091)	0.068 (0.061)	0.113* (0.048)	0.002 (0.004)
Endline	−0.073 (0.062)	−0.068 (0.047)	0.028 (0.068)	−0.060 (0.046)	−0.041 (0.035)	−0.005^{+} (0.003)
Constant	0.614**(0.097)	0.573**(0.073)	0.682**(0.106)	0.746**(0.071)	0.637**(0.055)	0.998**(0.004)
Mean	0.773	0.656	0.785	0.744	0.763	0.996
Observations	40	40	40	40	40	40
R-squared	0.730	0.616	0.619	0.860	0.694	0.766
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note. Indicators were categorized with overlap allowed. (1) Attitude: Caregiver's attitude toward children. (5) Treated: Treated indicators. (6) Nontreated: Nontreated indicators. $^{+}$: $p < 0.10$. *: $p < 0.05$. **: $p < 0.01$. Standard errors are shown in parentheses.

Table A9

Verification of regression to the mean.

Variables	Selected items
Treat × Endline	0.037 (0.042)
Endline	0.000 (0.028)
Constant	1.000** (0.048)
Observations	40
R-squared	0.871
Controls	Yes

Note. Selected items were those that should have undergone intervention under the feedback design within the control group, excluding the actually intervened items within the treatment group. Standard errors are shown in parentheses. **: $p < 0.01$.

Table A10

Results of verification check using robust standard errors, accounting for heteroscedasticity and clustering of errors within facilities.

	Total	Space and Furnishings	Personal Care Routines	Language and Literacy Learning	Activities	Interaction	Program Structure
Treat×Endline	0.527	0.455	0.553	0.345	0.741	0.489	0.135
Homoskedastic SE	(0.183)	(0.330)	(0.404)	(0.344)	(0.229)	(0.310)	(0.530)
Cluster Robust SE	(0.265)	(0.446)	(0.586)	(0.514)	(0.334)	(0.430)	(0.794)

	Caregiver's attitude toward children	Objects related	Time use	Other	Treated items	Nontreated items
Treat×Endline	0.125	0.156**	0.121	0.029	0.113**	0.002
Homoskedastic SE	(0.081)	(0.056)	(0.078)	(0.055)	(0.047)	(0.004)
Cluster Robust SE	(0.114)	(0.079)	(0.106)	(0.074)	(0.064)	(0.006)

A11

Summary of the analysis for the Teacher belief scale.

Before the ECERS-3 observations in 2020, a survey, including the Teacher belief scale (Stipek & Byler, 1997), was conducted with elementary school teachers and child caregivers in the same city where the present study was conducted. The Teacher Belief Scale is a questionnaire concerning teachers' beliefs about appropriate education for young children. In the original version, teachers are asked to answer the degree to which they agree with 31 statements on a 5-point Likert scale. The items are categorized into two beliefs: the "basic-skills orientation belief," including items indicating that the teacher believed in the effectiveness of formal, highly structured instruction of basic skills, and the "child-centered orientation belief" including items indicating the developmentally appropriate practices, emphasizing children's self-initiation, control, and exploration of concrete objects (Stiepek & Byler, 1997). Excluding three items that did not fit the context of Japanese language acquisition or Japanese elementary school or kindergarten, 22 items (12 items for basic-skills orientation belief and 10 items for child-centered orientation belief) were used in our survey. The average value of the item scores included in each belief was used in the analysis (Basic-skills orientation belief: $\alpha = .72$; Child-centered orientation belief: $\alpha = .61$). The child-centered orientation belief score was significantly higher than the basic-skills orientation belief score for both child caregivers and primary school teachers (child caregivers: $t(350) = 17.309$, $p < 0.001$, $d = 1.548$, 95% CI [1.383, 1.713]; primary school teachers: $t(128) = 6.204$, $p < 0.001$, $d = 0.926$, 95% CI [0.674, 1.177]). The child-centered orientation belief score was significantly negatively correlated with the basic-skills orientation belief score for both child caregivers and primary school teachers (child caregivers: $r(349) = -.355$, $p < 0.001$; primary school teachers: $r(127) = -.448$, $p < .001$). The within-individual difference between the two beliefs was

significantly greater for child caregivers than primary school teachers, suggesting that child caregivers have a stronger child-centered orientation than primary school teachers (Mean difference: child caregivers = 0.637, primary teachers = 0.444; Welch two sample t-test: t (199.57) = 2.3998, $p < 0.05$, d = 0.267, 95% CI [0.063, 0.470]).

Descriptive statistics for the Teacher Belief Scale.

	Basic skills orientation		Child-centered orientation	
	N	Mean (SD)	N	Mean (SD)
Total	498	3.006 (0.471)	509	3.600 (0.405)
Child caregivers	366	2.957 (0.449)	370	3.607 (0.390)
Primary school teachers	132	3.143 (0.505)	139	3.582 (0.444)

Note. The number of observations is varied because some items have missing values. Cases with multiple responses to an item were treated as missing values.