



## Examining the effects of an infant-toddler school readiness intervention in center- and family-based programs: Are results generalizable?

Dorthe Bleses<sup>a,b,c,d,\*</sup>, Peter Jensen<sup>a,b,c,d</sup>, Anders Højen<sup>a,b,c,d</sup>, Marinka M. Willemsen<sup>a,b,c,d</sup>, Pauline Slot<sup>e</sup>, Laura M. Justice<sup>f</sup>

<sup>a</sup> School of Communication and Culture, Aarhus University, 8000, Aarhus C, Denmark

<sup>b</sup> TrygFondens' Centre for Child Research, Aarhus University, 8000, Aarhus C, Denmark

<sup>c</sup> Department of Economics and Business Economics, Aarhus University, 8000, Aarhus C, Denmark

<sup>d</sup> Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark

<sup>e</sup> Department of Child, Family, and Education Studies, Utrecht University, the Netherlands

<sup>f</sup> Department of Educational Studies, The Ohio State University, USA

### ARTICLE INFO

#### Keywords:

School readiness intervention  
Toddler classrooms  
Family-based and center-based programs  
Language, math and social-emotional child outcomes  
Educational practice outcomes

### ABSTRACT

Infants and toddlers frequently participate in either center- or family-based childcare programs. However, little is known about the efficacy of early learning interventions introduced in these two types of programs, in particular family-based programs. The present work builds upon findings of a recent experimental trial demonstrating that a 20-week infant-toddler intervention supporting center- and family-based teachers to be more explicit and intentional in their interactions had a significantly positive effect on targeted child outcomes. In this follow-up paper, we conducted secondary analyses exploring effects of the intervention across the two contexts, center- and family-based programs. Analyses showed that the social validity of the intervention was generally high in both settings, but even higher in family-based than center-based programs. Findings also showed that teachers in both types of programs implemented the intervention at a satisfactory level, but family-based teachers tended to implement more small-group activities and had more conversations with individual children. There were no differential impacts on child outcomes across the two contexts, except for an overall significant spill-over effect on the outcome of empathy within center-based care. Finally, we found that the intervention had positive effects on teachers' use of counting and math activities in both types of programs.

Early childhood education (ECE) is widely considered an important means to foster young children's development in domains such as language and literacy, math, and social-emotional skills, with potential value for mitigating the school-readiness gap between children from less-advantaged backgrounds and their advantaged peers (see Burchinal et al., 2015 for an overview). Enrollment in high-quality ECE programs has large beneficial effects on these child outcomes as well as future academic achievement, employment, and earnings (e.g., Duncan & Magnuson, 2013; Duncan et al., 2022; Philips et al., 2017). However, much of this work is focused on preschool programs, with relatively less focus on infant-toddler ECE programs (e.g., Ackerman, 2021; Greenwood et al., 2020; Walker et al., 2020).

Currently, about one-third (36 %) of infants and toddlers in OECD countries are enrolled in either center-based or organized family-based programs provided by a professional teacher (herein, we use the term

teacher to refer to staff who work in the two types of programs), although there is large variation among countries (1 to 65 % enrollment; OECD, 2021). Family-based programs, as compared to center-based programs, tend to be characterized by an informal home-based setting with oftentimes less qualified staff and lower teacher-child ratios (e.g., Bigras, 2010). Family-based programs account for up to one third of ECE-enrolled children in OECD countries (e.g., Ofsted, 2020, Statistics Denmark, 2020; Vandenberg et al., 2021).

Compared to ECE programs targeting preschool-aged children, studies find generally lower and more heterogeneous quality of ECE programs for infants and toddlers in the United States (U.S.) (Burchinal et al., 2015). Moreover, when directly compared with center-based programs, family-based programs appear to have the lowest quality in various OECD countries (Belgium, Canada, Germany and United Kingdom and the U.S.), although results are mixed across studies

\* Corresponding author at: TrygFondens Centre for Child Research, School of Communication and Culture, Aarhus University, 8000, Aarhus C, Denmark.  
E-mail address: [bleses@au.dk](mailto:bleses@au.dk) (D. Bleses).

<https://doi.org/10.1016/j.ecresq.2024.01.001>

Received 17 February 2022; Received in revised form 4 January 2024; Accepted 4 January 2024

Available online 17 January 2024

0885-2006/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(Bigras, 2010; Eckhardt & Egert, 2020; Vandenbroeck et al., 2021). Such findings highlight the need to improve the quality of both center-based and family-based infant-toddler programs, although few studies have attempted to experimentally test the effectiveness of interventions aimed at doing so (cf. Bleses et al., 2021; Helmerhorst et al., 2017; Werner et al., 2018). Note that “school readiness” for toddlers should be thought of as skills that are foundational for developing school readiness throughout their lives before entering school (Lally, 2010).

To contribute to this important research base, in this paper we present the results of secondary analyses of a school-readiness intervention with a particular emphasis on vocabulary development implemented in Denmark’s universal infant-toddler ECE programs. The aim of the present work was to explicitly examine differences and similarities between center- and family-based programs in relation to social validity and implementation fidelity of the intervention “We learn together”, and to compare intervention effects on child and teacher outcomes across the two types of programs. To our knowledge, this study is the only one to have directly compared the effects of the *same* intervention implemented in both center-based and family-based child-care programs.

## 1. Infant-toddler ECE programs: what is the overall quality of the learning environment?

A growing body of literature has examined the relations between the type of childcare programs and children’s school readiness outcomes in kindergarten. Few studies have directly compared center-based programs with formal (i.e., regulated) family-based programs. The available evidence suggests that center-based programs have a significantly positive impact on measures of school readiness whereas family-based programs have no or even negative impact (e.g., Ackerman et al., 2021; Slot, 2018).

The extent to which findings like these are related to the quality of the learning environment in the two types of ECE programs has been addressed in several studies where the ECE quality model is based on assessments of the structural and process quality of the programs (Bigras, 2010; Bumgarner & Brooks-Gunn, 2015; Côté et al., 2013; Eckhardt & Egert, 2020; Slot, 2018). Structural quality concerns the features that support process quality including teacher-child ratio, teacher qualifications, and program infrastructure, whereas process quality comprises children’s day-to-day experiences, teacher language practices and learning opportunities, and interactions with their teachers and peers (Slot, 2018). Some studies have found that center-based programs have higher process quality than family-based programs (e.g., Bigras, 2010; Bumgarner & Brooks-Gunn, 2015), whereas other studies have shown that the quality of family-based programs is at comparable levels with center-based care or even higher (Côté et al., 2013; Eckhardt & Egert, 2020; Vandenbroeck et al., 2021). A cross-cultural meta-analysis aggregated the process quality of family-based programs based on 30 (mainly U.S.) studies, whereby quality was assessed using the same instruments (the Family Day Care Rating Scale [FDCRS, Harms & Clifford, 1989] and the Family Child Care Environment Rating Scale (Harms et al., 2007)). The meta-analysis found medium quality scores at levels comparable with center-based programs (Eckhardt & Egert, 2020). In contrast, a later systematic review which included 25 studies from the U. S. and 11 studies that were based in Europe revealed that the relations between enrollment in infant-toddler ECE programs and child outcomes are mixed and appear to be partly associated with structural and process quality of the program in which children are enrolled (Ackerman et al., 2021).

However, recent research indicates that assessments of the structural and process quality of ECE programs only to a limited extent correlates with child outcomes (see, e.g., Burchinal, 2018; McDoniel et al., 2022) indicating that improvements in these aspects of the learning environment may not have impact on children’s outcomes (for conflicting evidence see Von Suchodoletz et al., 2023). Rather, research using research instruments that allow for a more finegrained perspective on

educational practices than the global models of quality, has demonstrated that teachers’ practices in the classroom are important predictors of children’s outcomes and that such practices can be improved through interventions implemented in ECE programs (Zauche et al., 2017; Walker et al., 2020). A recent meta-analysis by Hadley et al. (2022) including 30 studies and 539 correlations identified two dimensions of effective teacher language practices comprising an Emergent Academic Language dimension (abstract conceptual talk with a greater variety of word types) and a Bridge Language dimension (scaffolding practices including questions and elicitations and responsive talk). Such practices have been linked to preschoolers’ language outcomes (see, e.g., Burchinal et al., 2021; Hadley et al., 2023).

Fewer studies have identified associations between teacher language practices and toddlers’ outcomes. Bratsch-Hines et al. (2020) examined teacher practices in ECE settings targeting infant and toddlers and relations to immediate outcomes and later associations with language, social, and academic skills measured at 36 months and in kindergarten in a sample of 1055 children. Findings indicated that children who experienced more positive caregiver-child verbal interactions had higher language skills at 36 months, which indirectly led to higher kindergarten academic (including language) and social skills. A similar result was found by Mitsven et al. (2022) in a small study using a slightly younger sample of 2.5–3.5-year-old children with or without a hearing loss. A measure of adult vocalizations (indexed by the number of speech-related vocalizations, e.g., words, utterances, and conversational turns) obtained from child-worn audio recorders over 34 observations spanning a school year predicted child vocalizations (pre-linguistic sounds, babbling, protophones and full word production).

Improving specific teacher language practices in activities and daily routines therefore appear to be a promising avenue to promote early language skills, though evidence from toddler classrooms is more limited. Moreover, it is not clear whether the lack of impact of family-based programs on child outcomes is related to differences in global quality of the two types of programs or teacher language practices.

## 2. Key characteristics of interventions promoting effective educational practices that can improve the learning environment of ECE programs

The increase in the development and testing of interventions in high-quality studies to support children’s learning experiences has been dramatic, and across studies a set of effective characteristics have been identified (Alves et al., 2023; Blinkoff et al., 2023; Burchinal & Farran, 2020; Duncan et al., 2022; Hadley et al., 2022; Hadley et al., 2023; Walker et al., 2020; Weiland et al., 2018). Effective interventions often support (1) an explicit focus on strengthening early skill development (e.g., using a sequence and scope); (2) teachers’ use of complex language and scaffolding strategies in consistent and responsive conversations; (3) use of guided play and interesting activities that engage children and support content; (4) a positive and emotionally supportive atmosphere; (5) use of data about intervention implementation; and (6) professional development of staff.

Few studies to date have examined the effectiveness of interventions aimed at improving the quality of infant-toddler ECE programs that include all the effective intervention elements listed above. In particular, very few interventions have had a specific focus on skill development in infant-toddler childcares (e.g., Bleses et al., 2020; Landry et al., 2014). There is a need to promote early skill development in the infant and toddlerhood years, as early language, math and social-emotional skills are associated with later skills in school-aged children (Duncan et al., 2007; Dale et al., 2023). Providing infant-toddler teachers with a systematic scope and sequence that clearly identifies specific skills to target represents a potentially important means to support skill development.

Landry et al. (2014) evaluated an infant-toddler curriculum that targeted social-emotional skills via an explicit scope and sequence whereas language was targeted through teacher responsive interactions.

The study found positive impacts on targeted teacher outcomes (e.g., richer language input and more engaging learning activities) and child social-emotional skills but not for language skills. Bleses et al. (2020) evaluated the effects of a 20-week school readiness intervention providing teachers with a sequence and scope for language and math development and supportive tools to be more explicit and intentional in their interactions with children. This study found positive, mainly medium- to large-sized effects on child outcomes in both language and math. Thus, there is some evidence that providing teachers with a sequence and scope is also an effective element in interventions targeting infant-toddler classrooms in center-based childcares.

Interventions in infant-toddler classrooms in center-based childcares targeting both types of teacher language practices (Emergent Academic Language dimension and Bridge Language dimension) has shown some effect on child outcomes (Bleses et al., 2020; Landry et al., 2014). Additionally, a few studies have tested interventions aiming at improving such teacher language practices specifically within family-based programs, but the conclusions that can be drawn are limited due to small sample sizes, lack of comparisons with other program types, and/or lack of measures of quality and child outcomes (Koh & Neuman, 2009; Merz et al., 2016). For instance, one small-sized intervention with a pre-post design conducted by Koh and Neuman (2009) investigated the effects of professional development on scaffolding practices. Teachers in family-based programs participated either in a 15-week course on language and literacy ( $N = 26$ ), this course plus coaching ( $N = 32$ ), or business as usual ( $N = 49$ ). The teachers who received the course (e.g., support for learning and scaffolding strategies, physical environment) and coaching showed higher performance on educational practices than the control and course-only group but the effects on child outcomes were not investigated. A more recent randomized controlled trial (RCT) was conducted by Merz and co-workers (2016) in which teachers in U.S. family-based programs received a 20-week course on emotional and educational responsiveness ( $N = 62$ ). The teachers participating in the course outperformed their colleagues in the control group concerning responsive behaviors, and the youngest children in the treatment group showed improvements in inhibition and attention. Finally, Moreno and colleagues (2015) evaluated the effectiveness of two professional development interventions aimed at improving the quality of care provided by teachers in both center- and family-based infant-toddler programs. Findings showed that a combination of coursework and coaching improved teacher practices related to the quality of interactions. This study did also not include child outcome measures.

In summary, study findings indicate that teacher-focused interventions can improve some teachers' language practices, but these studies did not analyze impacts across different types of ECE programs, nor did they evaluate effects on child outcomes. It is therefore unclear whether higher levels of effective teacher practices translated into improved child outcomes. There is therefore an urgent need to develop and test the effectiveness of interventions targeting teachers and children in both types of programs, particularly in universal programs with high levels of enrollment of infants and toddlers, as is the case in Denmark.

### 2.1. The Danish universal ECE system

The Danish childcare system features nearly universal enrollment of children, and all municipalities are obliged to ensure availability of programs (Gupta & Simonsen, 2010). Currently, three of four children are enrolled in ECE programs before they turn three years and are on average about ten months old at program entry (Ministry for Children and Social Affairs, 2018a).

The universal ECE system in Denmark is characterized by a high level of public expenditure compared to other countries and by comparatively high structural quality (Nøhr et al., 2012). Most children are enrolled in center-based childcare (71 %, Statistics Denmark, 2020), and the

majority serve children in the entire age range from 0 to 5 years (Ministry for Children and Social Affairs, 2018a). While three- to five-year-olds are all enrolled in center-based programs, one- and two-year-olds are in either center-based programs with typically eight to 12 children or in family-based programs (Nøhr et al., 2012). Center-based programs vary in size (range 11–80 children), with most centers enrolling an average of 40 to 50 children; some centers serve the entire age range 1–5, but with separate classrooms for infant-toddlers (Nøhr et al., 2012). The teacher-child ratio in centers is relatively low, with one teacher per about 3 children (Statistics Denmark, 2020). Approximately 60 % of teachers have a 3.5-year pedagogical bachelor's degree (but with some local variation, Ministry of Children and Social Affairs, 2018b).

Family-based programs for infant-toddlers serve up to five children cared for by a teacher in his/her own home. Individual teachers are organized in larger groups (family-based childcare groups) with typically five to six teachers supervised by a family-based childcare teacher (often with a formal BA education) who is the daily leader of the family-based teachers in the group. The actual organization of the family-based childcare units varies substantially across municipalities. A family-based childcare teacher has on average 3.3 full-time children, but with some local variation (Statistics Denmark, 2020). There are no formal education requirements for family-based teachers in Denmark. About 8 % of family-based teachers completed a three-year pedagogical assistant education, about one-half attended vocational education unrelated to the ECE field, and the remainder did not complete any education after secondary school (Statistics Denmark; [www.statistikbanken.dk/BOERN1](http://www.statistikbanken.dk/BOERN1)). The socioeconomic composition of both center-based and family-based programs is mainly determined by the composition of the neighborhood.

All but three of the 98 Danish municipalities offer both types of programs for infant-toddlers. Parents can, in principle, choose which type they want for their child. But even when both types of programs are offered, the nearest center may in some cases be too remote for parents to choose that option. In addition, each option may not be available at all times, and consequently, parents may not have a real choice if they want care for their children here and now in order for both parents to be able to work. Therefore, there are likely to be certain biases with respect to the types of family who choose each type of program (or have a real choice altogether).

A large-scale study ( $N = 30,444$ ) of associations between the structural quality of ECE childcare programs targeting preschool-aged children and educational achievement at the end of primary school has shown that a higher staff-per-child ratio, a higher share of male staff, and a higher share of staff with formal preschool teacher training was associated with significant improvements in children's test results in Danish (Bauchmüller et al., 2014). Similarly, another large-scale Danish study of the whole population of non-immigrant Danish toddlers born in 1994–1995 found that toddlers who were enrolled in center-based care at age two years had significantly higher school leaving grades and more often attended the academic track in high school as compared to the toddlers in family-based programs (Gupta & Simonsen, 2016).

The quality of the learning environment in the Danish ECE system has been less investigated. The ECE area is regulated by the "childcare legislation" [Dagtilbudsloven] which emphasizes an enhancement of the ECE learning environment throughout the day by focusing on topics aligned to six curriculum themes (e.g., personal and social-emotional development, language, math and science) and strategies that support the involvement of all children. Even though this learning curriculum is used in both types of programs, there has been very limited focus on reinforcing the broad learning curriculum in family-based programs compared to center-based care.

The quality of Danish infant-toddler ECE programs has been assessed to a limited extent. A recent study showed comparable quality of the learning environment between center-based infant-toddler and preschool programs using a global model of quality (Bleses et al., 2020). A

recent national observation study, which includes infant-toddler center-based as well as family-based childcares, demonstrates that only 10 % of center-based childcares and 8 % of family-based childcares have a “good” quality, between 49 % (center-based) and 39 % (family-based) of childcares have ‘sufficient’ quality, whereas between 38 % (center-based) and 46 % (family-based) of childcares have ‘inadequate’ quality (Lindeberg et al., 2023).

## 2.2. Study aims

In this follow-up paper, we conducted secondary analyses of an RCT study of a school readiness intervention (“We learn together”, previously “Play and Learn”) implemented in both center- and family-based infant-toddler ECE programs in Denmark. The core components of the intervention include a scope and sequence of instruction targeting language and math that is laid out in a 20-week curriculum. To address this scope, teachers are asked to develop activities of their own choice to address specific skills within a theme-based instructional framework. Also, teachers are provided with supportive tools to be more explicit and intentional in their interactions with children, as detailed later in this report. In a previous paper (Bleses et al., 2021) we tested whether we were able to replicate the effects of an earlier-stage assessment of the curriculum with additional implementation support (Bleses et al., 2020). The main results of the study, based on a heterogeneous sample of 2170 one- and two-year old children, demonstrated overall positive and (according to Kraft, 2020) medium- to large-sized effects on targeted language and math skills ( $d = 0.10$ – $0.59$ ). However, a non-significant result was found for children’s language use ( $d = 0.03$ ). We also explored if type of program moderated the intervention effects, finding no differential effects of the intervention for children in center-based vs. family-based ECE programs.

In the current study, we build upon the prior results by conducting secondary analyses of the data presented in the original paper and by analyzing additional data and focus explicitly on examining social validity, intervention implementation, and child outcomes between center-based and family-based programs not presented in the original paper. Specifically, we addressed three questions: (1) What is the social validity and implementation fidelity of the “We learn together” intervention across the two types of programs, center-based vs. family-based? (2) Are there any differences in intervention effects on children’s language, math, and social-emotional outcomes across the two types of programs? And 3) Does participation in the intervention result in effects on targeted and non-targeted educational activities in the two types of programs?

## 3. Methods

The design, procedures, and participants of the study are reported in detail in the main study manuscript (Bleses et al., 2021), and we refer readers to that report. Here, we present details of the study most pertinent to the present study questions.

### 3.1. Design and participants

The “We learn together” intervention was experimentally evaluated using a cluster-randomized trial running for two consecutive periods. All results reported in the current paper are based on data from the first period. All 98 municipalities in Denmark were invited to participate and 13 accepted the invitation. The participating center- and family-based childcare units were recruited from these 13 different municipalities. For seven of these municipalities, all their childcare units were enrolled in the study. For the remaining six municipalities, 10 childcare units were randomly selected from each municipality. This resulted in the enrollment of a total of 255 childcare units; 111 were center-based and 144 were family-based childcare groups (i.e., groups of family-based childcare teachers who meet regularly). All children in the participating childcare units were enrolled in the study.

All childcare units were randomly allocated to either the treatment group in which “We learn together” was implemented or to the control group (business as usual). Randomization occurred at the level of center- and family-based childcare groups, respectively and was done prior to the start of the first period. The randomization was done by generating a sequence of random numbers in Stata (using *runiform*) and then applying these to a list of the participating childcares, thereby randomly assigning them to either the treatment group or the control group.

The number of participants was determined using a priori power calculations such that we would be able to detect an effect size of  $d = 0.2$  with a power of 80 % and a standard significance level of 5 %. Data for children and teachers were collected between August 2016 and March 2017. We experienced attrition among the participants, both at the cluster and individual level, as shown in the flow chart in Fig. 1. The cluster-level attrition meant that 6 childcares did not provide any post-test data (3 in the treatment group and 3 in the control group; 2 % and 4 % respectively). One childcare closed during the intervention period, whereas the other 5 did not collect post-test data for unknown reasons. At the individual level, pre-test data were available for 3588 children, whereas post-test data were available for 2170 children. No differential attrition was found, since the attrition rates were almost identical in the treatment group and the control group (39 % and 40 %, respectively; not significantly different). The main reason for children to drop out of the study before the post-test was that they left the childcare to transition to preschool and in fewer cases because of residential mobility. Families with young children are often in need of different (and larger) housing arrangements and when the families change their residential location, they typically move their children to a childcare near their new home. When we assessed the attrition rates of the study using the standard of the U.S. Department of Education, What Works Clearinghouse (Clearinghouse, 2017), they fell in the tolerable region both at cluster level and at individual level, even under cautious assumptions (see Bleses et al., 2021, for more information).

In Table 1, the baseline characteristics of participating childcares, teachers, and children per type of program are shown. We obtained information from administrative registers via the Danish civil registration system on children and family background using Statistics Denmark, to include parents’ education, immigrant status, country of origin, income, and employment status, among others. Balancing tests of baseline characteristics of the teacher, children, and parents showed that these were distributed evenly across the control and treatment groups in each of the two types of programs (see Table 1). The final analytic sample for the study of children’s outcomes consisted of 2170 children, corresponding to all children aged 18 months or over who had both pre-test and post-test scores for a given outcome measure. In our sample, about twice as many children were in center-based programs than family-based programs, which reflects the proportion of children in each type of program in the population.

### 3.2. Procedure

#### 3.2.1. Pilot study to determine the quality in family-based childcares using a global model of quality

In order to assess the extent to which the intervention, first developed for center-based programs, was adequate for family-based childcare as well, as part of the current study we conducted a pilot study to estimate the quality of the learning environment from self-selected family-based childcares which would enable preliminary quality comparisons between center-based (published in Bleses et al., 2020) and family-based programs. We invited family-based childcares from the 13 included municipalities to participate in the pilot study and 45 family-based childcares volunteered. Observations were conducted prior to the intervention study. The result of this comparison is presented in Table 2. In both center- and family-based childcares classroom quality was measured by CLASS-Toddler (LaParo et al., 2012), and teacher-child interactions were rated on a 7-point scale. Scores ranging



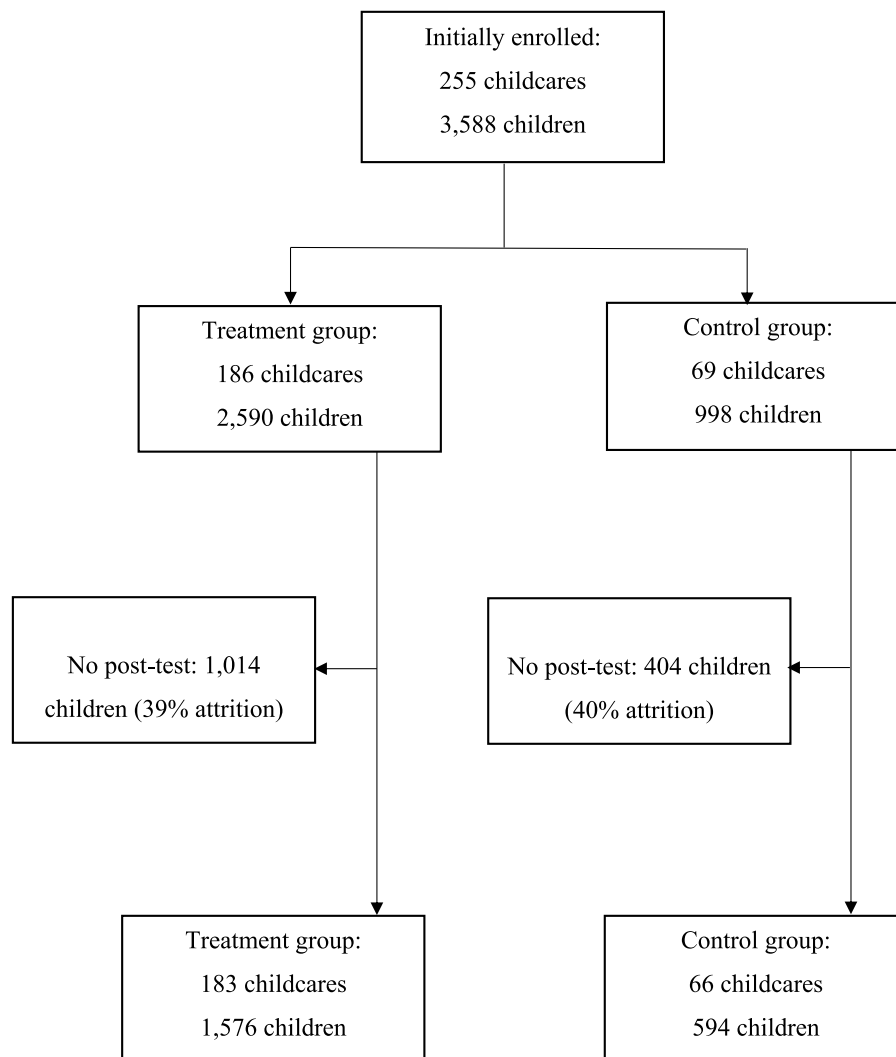


Fig. 1. Flow chart of the randomization and analysis process.

from 1 to 2 were considered low, scores ranging from 3 to 5 were considered in the midrange, and scores from 6 to 7 were considered high. CLASS domain scores were close to or in the same quality range for center- and family-based childcares. For both types of programs, teachers were found to display medium to high responsiveness to children's emotional and behavioral needs. Concerning teachers' ability to facilitate children's learning and development, the quality of center-based childcare fell barely outside the midrange of quality but resembled the quality of family-based childcare that fell just in the midrange. Moreover, classroom quality varied more between family-based programs than between center-based programs. Only teachers' behavior guidance clearly varied more between center-based than between family-based programs. Although the pilot data of family-based programs must be interpreted with caution (mainly because teachers from family-based programs self-selected into the pilot study but also because the sample size is smaller than that in center-based childcare), classroom quality appears to be very comparable between Danish center- and family-based programs and indicate that no substantial modification of the intervention was needed.

### 3.2.2. Intervention

All center-based and family-based programs in the treatment group implemented a researcher-developed 20-week "We learn together" intervention. The intervention is theme-based and includes 5 content areas which are relevant and engaging for toddlers (e.g., My Family; Out

in the World). The intervention is described in detail in [Bleses et al. \(2021\)](#), and here we focus on how the core elements of the intervention were adapted to the different programs. The intervention captures several of the different components which have been identified as particularly promising in recent reviews including playful learning and effective teacher practices (e.g., [Blinkoff et al., 2023](#); [Hadley et al., 2022](#); [Joo et al., 2020](#); [Weiland et al., 2018](#)), as follows.

First, as children's language skills are developed early compared to other skills ([Cates et al., 2012](#)), promoting language skills play a central role in the intervention. All childcares implemented a two-fold scope of instruction targeting language (general vocabulary, language use), and math (math vocabulary, numeracy skills) that are sequenced over the 20-week curriculum period (effectively implemented over a 7-month period because of various holidays). Objectives for vocabulary included, for instance, the following categories: (1) *general* vocabulary: thematic words (typical words for objects, events and actions, e.g., "beach", "wet"), words for feelings (e.g., "happy", "angry"), time (e.g., "before", "now") and space ("down", "under"); and (2) *math* vocabulary: words for numbers, shapes (e.g., "round", "square"), sizes (e.g., "big", "short"), and patterns (e.g., "dots", "stripes"). To support implementation, each classroom in center-based childcare and each individual family-based childcare unit received five books, small and large picture cards, and posters depicting all target words. The sequence and scope encouraged teachers to use more abstract conceptual talk with a greater variety of word types (the Emergent Academic Language dimension, cf.

**Table 1**  
Baseline characteristics of childcares, children, and teachers (by group) and balancing tests, by type of program.

	Center-based program			Family-based program		
	Control	Treatment	Balancing tests	Control	Treatment	Balancing tests
No. of centers / groups	29	81		37	102	
No. of classrooms / units	68	173		112	326	
No. of children	371	982		223	594	
No. of teachers	75	230		92	299	
	Mean	Mean	<i>p</i> -value	Mean	Mean	<i>p</i> -value
<i>Teacher characteristics</i>						
% female	100.0	97.4	0.16	95.6	99.3	0.01*
Age (mean in years)	45.1	44.6	0.69	50.2	49.3	0.35
% > 10 years of experience	61.3	63.5	0.74	68.5	68.2	0.96
% BA degree	78.7	74.9	0.41	4.3	4.3	1.00
<i>Child characteristics</i>						
% girls	46.9	51.5	0.15	48.9	48.8	0.99
Age (mean in months)	23.3	23.6	0.32	22.7	23.1	0.10
% Danish origin	87.6	85.7	0.65	93.7	94.4	0.75
% Western origin	3.2	3.3	0.99	1.8	2.7	0.48
% non-Western origin	9.2	11.0	0.61	4.5	2.9	0.32
% Maternal education: low	16.7	16.5	0.94	13.5	16.3	0.35
% Maternal education: low-mid	30.7	32.2	0.59	35.0	42.3	0.07
% Maternal education: high-mid	31.5	30.3	0.74	33.2	28.8	0.24
% Maternal education: high	17.3	16.3	0.83	13.9	10.1	0.15
% Maternal education: missing	3.8	4.7	0.65	4.5	2.5	0.40
% Paternal education: low	16.4	18.7	0.36	16.1	19.7	0.25
% Paternal education: low-mid	42.0	39.9	0.54	57.8	53.0	0.26
% Paternal education: high-mid	21.0	19.7	0.57	13.9	17.3	0.26
% Paternal education: high	13.7	14.8	0.79	9.0	6.6	0.30
% Paternal education: missing	6.7	6.9	0.91	3.1	3.4	0.89
% Parents married or cohabiting	83.2	84.6	0.58	89.6	88.0	0.52
% Parents without employment	14.8	12.7	0.53	8.5	8.8	0.93
% Owner-occupied housing	56.9	56.8	0.99	73.9	72.2	0.70

Notes: Data for children are only based on children aged 18 months or over at the time of the pre-test and with both pre-test and post-test assessments. Data for teachers are only based on teachers who answered both the pre-test and the post-test questionnaire. Balancing tests are for equality of means in the control and treatment group. For child characteristics, the *p*-values are from regressions of row variables on a treatment indicator variable with standard errors adjusted for clustering at the daycare center/group level.

\* *p* < 0.05.

**Table 2**  
Descriptive information for the CLASS Toddler for Danish center-based (*N* = 80) and family-based programs (*N* = 45).

	Center-based program			Family-based program		
	Mean	SD	Range	Mean	SD	Range
<i>Emotional and Behavioral Support</i>	5.79	0.45	4.7–6.9	5.82	0.51	4–6.8
Positive Climate	6.01	0.52	4.9–7	6.18	0.50	5–7
Negative Climate*	1.24	0.29	5.8–7	1.21	0.37	1–3
Teacher Sensitivity	5.50	0.59	4.1–6.8	5.94	0.74	3–7
Regard for Child	5.11	0.64	3.4–7	4.04	0.86	2–6
<i>Perspective Engaged Support for Learning</i>	2.96	0.49	1.9–4.2	3.15	0.67	1.7–4.6
Behavior Guidance	5.57	0.75	3.8–7	6.14	0.54	5–7
Facilitation of Learning and Development	3.55	0.57	2.4–5	4.02	0.97	2–6
Quality of Feedback	2.26	0.63	1–4.5	2.25	0.55	1–4
Language Modeling	3.09	0.61	1.9–4.6	3.18	0.71	1.7–4.7

Note. The results of family-based care were part of a pilot study.

\* Reversed scale scores.

Hadley et al., 2022)). To support the targeted skills, teachers were asked to develop playful activities of their own choice and to use teacher language practices to enrich and differentiate their interactions with individual children, including contingently responding to children’s comments and engaging children in extended conversations (the Bridge Language dimension, cf. Hadley et al., 2022). Implementation materials like posters and videos that endorse the use of specific teacher language practices supporting enriched conversations and a weekly planner offering a theme-based instructional framework by which to plan these

activities in alignment with the week objectives (see Bleses et al., 2021, for more details).

Second, the suggested dosage of participation for individual children per week was two large-group and two small-group activities and one individual conversation. The basic unit of planning and implementing the intervention in center-based programs is the classroom with individual teachers collaborating on the task. In family-based programs, however, both planning and implementation are distributed across two types of settings. Planning of the intervention was encouraged to take place in the family-based childcare groups (i.e., groups of teachers who meet in the same regular group). Coordinated planning was also necessary as large group activities are best implemented in family-based childcare groups, where children of multiple family-based teachers are gathered, whereas small-group activities and individual conversations are best implemented in the individual family-based childcare units.

Teachers in each classroom or family-based childcare unit completed implementation notes on a web-based logging tool and monitored child benefits from the intervention to support reflection on implementation. The implementation logs tracked the dosage of individual children in each classroom/family-based childcare, in the latter case child participation in the individual family-based childcare unit and in the family-based childcare groups. The self-reported implementation notes also tracked elements of adherence to the intervention (e.g., which educational activities did they use, and the extent to which the learning objectives were targeted, see also below). Teachers in family-based childcares reported on activities across the two types of settings.

### 3.2.3. Teacher professional development

It was obligatory for all staff in center-based and family-based childcares to take part in a 2-day professional development workshop which was held locally by educational consultants in the municipality

who had been on a 2-day central course held by the research team. On this workshop, the teachers were introduced to the main elements of the intervention. The workshop also included modeling of intervention strategies and the participants were provided with opportunities to practice the main content of the intervention (see Bleses et al., 2021 for more details).

### 3.3. Instruments

#### 3.3.1. Receptive vocabulary

Children's receptive vocabulary was tested with an adapted version of the English Computerized Comprehension Task (CCT; Friend & Keplinger, 2008; Friend, et al., 2012). In adapting CCT to Danish, we expanded the age range and developed three age-related difficulty levels for the ages 18–23, 24–29, and 30–35 months each consisting of items ranging from easy to hard for the age level. Moreover, the interface for the Danish adaptation was iPads. Scores represent the sum of correct answers. Each child was assessed on 39 items. The child sees two pictures representing a noun, verb, or an adjective on the iPad, a target and a foil, and the child is asked to touch the target word (e.g., touch the cat). Cronbach's alphas for each age group were: 0.88, 0.79 and 0.69 (see Bleses et al., 2021). Note that we aimed for similar age-dependent difficulty levels for the three versions.

#### 3.3.2. Productive vocabulary and language use

The domains were assessed based on the CDI-Educator, a brief teacher-based standardized checklist (Bleses et al., 2018). For vocabulary, scores represent the sum of correct answers. Each child was assessed with 70 items across 9 content categories (e.g., milk, find, soft). Language use includes five questions concerning the child's use of decontextualized language (e.g., whether the child at any time speaks about individuals who are not present or about something that will happen in the future). The three response categories were converted to points (from *not yet* [0] to *often* [2]) and summed up across the five questions. Test-retest correlations were acceptable (Vocabulary:  $r = 0.68$ ; Language use  $r = 0.54$ ). Internal reliability was high (vocabulary:  $\alpha = 0.98$ ; Language use:  $\alpha = 0.88$ ). Criterion validity was established through correlations with the Danish Receptive and Productive One-Word Picture Vocabulary Test ( $r = 0.43$  to  $0.65$ ). For more information about the instrument, see Bleses et al., 2018).

#### 3.3.3. Math language and numeracy

We administered a researcher-developed teacher-administered checklist reflecting the fact that no standardized measure for toddlers was available at that point in time. The checklist includes two dimensions of early math development, that is, children's comprehension and use of math language and numeracy. The two domains were primarily based on general work on children's math development, in particular numeracy development (Frye et al., 2013) and specific work on math-specific vocabulary laying the foundation for mathematical development (Purpura & Reid, 2016). Children were assessed with 10 items measuring early numeracy skills (e.g., saying random numbers, understanding of numbers, using number words) and 24 items measuring children's comprehension and use of math vocabulary (e.g., words for sizes, quantities, shapes, and space). The four response categories were converted to points (from *not yet* [0] to *always* [3]) and summed up across the five questions. For both constructs, all items had an item-total correlation higher than .50 and the internal reliability was high ( $\alpha = 0.95$  for both constructs). For more information about the instrument, see Bleses et al., 2020).

#### 3.3.4. Social-emotional skills

Teachers completed the Danish adaptation (Sjoe et al., 2019, 2020) of the standardized questionnaire, *Social-Emotional Assessment/Evaluation Measure (SEAM)–Research Edition* (Squires, 2014). SEAM assesses social-emotional skills through 10 benchmarks (e.g., empathy and

cooperation) which are rated on a scale from *not true* [1] to *very true* [4]. A high score indicates a positive aspect of social-emotional development. The ten benchmarks were converted to two overall indexes, the empathy index (the child's ability to communicate own feelings and to read and understand others' feelings), and the self-regulation & cooperation index (the child's ability to regulate and cooperate, and the child's adaptability). Reliability coefficients range from .82 to .91. For more information about the instrument, see Sjoe et al. (2019; 2020).

#### 3.3.5. Educational practices in the childcares

To assess to which extent the intervention affected the developmental and educational activities carried out in the childcares, we administered, in treatment as well as control classrooms, a structured questionnaire to teachers which was developed in the Netherlands (Slot et al., 2015) for preschool-aged children. However, the instrument was adapted slightly to be suitable in classrooms with infant-toddlers. The questionnaire includes a list of affective, play-supporting, self-regulation promoting and academic activities which were carefully chosen based on research into social-emotional and self-regulation development, and emergent academic skills, and organized into seven scales (Slot et al., 2015). We selected five scales which are considered indicators of educational quality and appropriate for infant-toddler classrooms. Each scale mix items relating to activities and different teacher practices. Two of those (language and math) are directly targeted in "We learn together" whereas the others are not. *Play activities* (7 items;  $\alpha = 0.80$ ) assesses the extent to which the teacher provides children with opportunities for free, self-managed play and occasionally enriches children's play, for example, "I let the children play without interfering". The response categories range from *not applicable* [0] to *strongly applicable* [4]. *Role play* (5 items;  $\alpha = 0.79$ ) assesses the extent to which the teacher stimulates cognitive distancing, symbolizing and pretend in children by modeling behavior and encouraging children to participate in symbolic and pretend play (e.g., I show children how to use an object for something else than intended, for instance driving a wooden block as if it is a car). The response categories range from *never* [0] to *very often* [4]. *Children's own planning* (10 items;  $\alpha = 0.80$ ) assesses the extent to which the teacher uses play, care routines and other activities to enhance children's behavioral self-regulation (e.g., when children have a conflict, I let them express their own opinion, so they better understand what the other thinks). The response categories range from *never* [0] to *very often* [4]. *Language activities* (15 items;  $\alpha = 0.87$ ) assesses the average frequency of activities and several forms of teacher language practices (e.g., having elaborate conversations about children's personal experiences, for instance what they did in the weekend). The response categories are *never* [0], *less than twice a month* [1], *twice or thrice a month* [2], *weekly* [4], *two to four times a week* [4], and 5 (*daily*). *Counting and math activities* (11 items;  $\alpha = 0.87$ ) assesses the average frequency of several math activities, for instance counting and sorting activities, and activities exploring different shapes (e.g., counting how many objects you have, for example counting till five and saying "I have five marbles"). The response categories are the same as for the language activities.

#### 3.3.6. Implementation fidelity

Child exposure to the intervention was used as a measure of implementation fidelity. We used the ongoing completion of weekly implementation logs on a web-based platform by the teachers in each classroom (one per classroom) or family-based childcare to track child exposure (number of times each child participated in large and small group activities per week and individual conversations).

Teachers also reported to what extent they adhered to the intervention, defined as the extent to which the intervention components were implemented (type of activities in large and small group settings; use of intervention material and activities, and addressing learning objectives).

### 3.3.7. Social validity

After the second period of the intervention, an evaluation questionnaire was administered to all center-based and family-based childcares ( $N = 252$ ) in the 13 municipalities. Management and staff in each childcare were requested to discuss a set of statements regarding participation in the study and about implementing the intervention. For instance, managers were asked to evaluate the statement “There was a positive attitude among the staff in relation to implementing the intervention” and staff were asked to evaluate the statement “The intervention was relevant and exciting”). Some items were related to aspects of social validity. We received responses from 98 center-based childcares and 55 family-based childcares as well as two responses that represented both center-based and family-based care. Two of the participating municipalities did not submit any evaluation forms. The staff were also offered to provide whatever feedback to the team they wanted.

### 3.4. Analytic strategy

To answer the first research question, we examined differences in social validity across childcare programs using descriptive statistics based on the questionnaire to management and staff. The differences in fidelity were similarly examined using descriptive statistics based on the measure of child-level intervention exposure.

To answer the second research question about differences in intervention effects on children’s outcomes, we estimated the effects of the intervention by using a value-added specification of a regression model:

$$y_i = X_i\beta + \delta I_i + \gamma y_i^{pre} + u_i,$$

where  $y_i$  is the post-test outcome measure for child  $i$  and  $I_i$  is an indicator variable for whether the child is enrolled in an intervention childcare. This makes  $\delta$  our main parameter of interest, since it represents the effect of the intervention compared with treatment as usual in the control group. To test whether the effects were different across the two types of programs, we included interaction terms between type of program and the intervention indicators (as well as estimating separate models for the two types of programs). We included the pre-test outcome measure  $y_i^{pre}$  to control for unobserved characteristics that might be correlated with prior achievement.  $X$  is a vector of explanatory variables that includes child and family characteristics (gender, age and ethnicity of the child, parental education level, parents’ marital status and employment status). These variables were included to control for individual differences that might influence the outcome measures and to increase the statistical precision of the estimates. We also included municipality fixed effects. These were included to control for potential municipality differences, which may arise due to the autonomy of the municipalities within the national framework and based on regulations. As mentioned above, the organization of childcares may vary across municipalities. Similarly, fees vary across municipalities. Outcomes were normalized with the standard deviations of the relevant pre-test for the control group, which means that estimates of  $\delta$  can be interpreted as effect sizes (Cohen’s  $d$ ). Since the randomization occurred at the childcare level, we adjusted the standard errors of the estimates to take account of this clustering. The cluster-adjusted standard errors also account for any other non-independence of children within childcares. In general, we analyzed the data at the child level with childcares as the clustering unit, since parts of the intervention happened at the childcare level (see Bleses et al., 2021, for more details).

For the center-based programs, the number of classrooms per center ranged from 1 to 8, with 54 % of the centers contributing multiple classrooms. Since many outcomes are based on teacher reports, it may be important to account for the hierarchical structure of the dataset with childcare centers and classrooms as separate levels. Hence, as a sensitivity analysis, we also used multilevel models (HLM) with 3 levels to investigate if it made any difference for the results to also take account of clustering at the level of classrooms/units.

As noted previously, the explanatory variables in the  $X$  vector were obtained from administrative registers at Statistics Denmark. Due to the high quality of these register data, we had a low level of missingness in the variables needed for the analysis. Primarily, parental education had missing values (as shown in Table 1), with 3.9 % of the values missing for maternal education and 5.5 % for paternal education. The other variables had less than 1 % missing values. In total, 8.5 % of the observations had one or more missing values. The main reason for parental education to be missing was that the education was completed outside Denmark. This would typically be related to other included variables such as origin and employment, so we treat these values as missing-at-random. In the statistical analysis, missing values were handled by using the ML method (FIML).

The third research question about differences in intervention effects on the educational practices across program was answered by estimating value-added models for the indices of the teachers’ activities, similar to those for child outcomes, but in this case controlling for the teacher’s age, gender, experience, and education. Again, differences across programs were tested by including interaction terms between type of program and the intervention indicators (as well as estimating separate models for the two types of programs). Due to the randomization at the childcare level, we also adjusted the standard errors of these estimates for this clustering.

## 4. Results

### 4.1. Preliminary findings

In Table 3, the mean scores of the child outcome measures are shown, separately for the control and the treatment group by program type. In general, the means showed a clear increase in the score from pre-test to the post-test, implying improved outcomes for the children over time. Across programs, scores tended to be slightly higher at pre-test for children in family-based programs, except for receptive vocabulary. Balancing tests showed that the pre-test outcome measures are balanced across control and treatment groups.

The summary statistics of the five indices for educational practices, reported by the teachers, are displayed in Table 4, separately for the control and the treatment group by programs. Four of the five indices had pre-test means in the upper half of the range of the indices, whereas the index for ‘counting and math activities’ had pre-test means in the lower half of the range of the index for both types of childcare. When comparing pre-test means for teachers in the two types of programs, we only found minor differences between teachers in family-based programs and teachers in center-based programs. There was a significant difference, however, for language activities (corresponding to an effect size of 0.2 of a standard deviation), indicating that teachers in center-based programs generally worked more intentionally with language supporting activities before implementing the “We learn together”. However, it is worth noting that considerable variation for all activities was observed. Thus, there were teachers in both center-based and family-based programs that reported that they did not use the practices described for each index at all.

### 4.2. Social validity and fidelity across childcare programs

To answer the first research question, we examined the social validity and fidelity of the “We learn together” intervention across the two types of programs (center-based programs,  $N = 98$ ; family-based programs,  $N = 55$ ). Overall, almost all teachers in both the center-based (94 %) and family-based programs (95 %) experienced that the intervention was relevant and exciting. There was generally a positive attitude among the staff toward implementation of the intervention (85 % in center-based care and 93 % in family-based care). Teachers in both center-based (97 %) and family-based programs (94 %) experienced during the implementation of the intervention that they were inspired by and



**Table 3**  
Pre-test and post-test scores for children’s outcomes, by type of program.

	N	Center-based program				N	Family-based program			
		Control		Treatment			Control		Treatment	
		Pre-test Mean (SD)	Post-test Mean (SD)	Pre-test Mean (SD)	Post-test Mean (SD)		Pre-test Mean (SD)	Post-test Mean (SD)	Pre-test Mean (SD)	Post-test Mean (SD)
Productive vocabulary (0–70)	1353	23.9 (17.6)	44.2 (17.4)	24.5 (17.6)	46.3 (18.6)	817	31.5 (21.1)	52.3 (17.0)	31.6 (19.1)	54.3 (15.9)
Receptive vocabulary (1–2) (0–39)	517	17.9 (10.0)	24.7 (6.2)	19.1 (9.2)	25.8 (6.5)	364	18.5 (9.6)	23.9 (7.3)	18.8 (9.8)	25.4 (6.4)
Receptive vocabulary (2–3) (0–39)	556	23.3 (6.8)	25.1 (4.9)	24.4 (5.7)	26.2 (4.8)	381	23.3 (7.1)	24.9 (5.6)	22.2 (7.5)	25.3 (5.6)
Language use (0–10)	1353	3.5 (3.0)	6.3 (3.0)	3.5 (2.8)	6.4 (3.0)	817	3.8 (2.9)	6.5 (2.7)	3.7 (2.7)	6.5 (2.9)
Math language (0–72)	1338	9.1 (9.9)	22.2 (14.7)	9.3 (9.4)	29.0 (15.6)	821	12.8 (11.7)	29.0 (15.3)	11.6 (10.5)	33.8 (15.3)
Numeracy (0–30)	1338	3.3 (4.6)	9.3 (6.7)	3.5 (4.8)	11.1 (7.2)	821	4.1 (5.5)	11.9 (7.1)	3.8 (4.7)	13.2 (6.8)
Empathy (0–18)	1255	11.8 (4.7)	14.4 (3.5)	11.6 (4.2)	14.9 (3.5)	784	12.9 (4.2)	15.7 (2.9)	12.5 (4.1)	15.7 (2.9)
Self-regulation & cooperation (0–12)	1255	7.7 (2.9)	9.1 (2.6)	7.8 (2.7)	9.3 (2.5)	784	8.2 (2.5)	9.4 (2.2)	7.9 (2.5)	9.5(2.3)

Notes: For receptive vocabulary, (1–2) indicates that the pre-test is from age-dependent test 1 (18–23 months) and the post-test is from age-dependent test 2 (24–29 months). Similarly, (2–3) indicates that the pre-test is from age-dependent test 2 and the post-test is from age-dependent test 3 (30–35 months). The other numbers in parentheses are the range of the measure.

**Table 4**  
Pre-test and post-test scores for educational practices, by type of program.

	N	Center-based program				N	Family-based program			
		Control		Treatment			Control		Treatment	
		Pre-test Mean (SD)	Post-test Mean (SD)	Pre-test Mean (SD)	Post-test Mean (SD)		Pre-test Mean (SD)	Post-test Mean (SD)	Pre-test Mean (SD)	Post-test Mean (SD)
Play activities (0–28)	305	18.9 (4.1)	19.8 (4.1)	18.8 (3.6)	20.1 (3.9)	391	18.6 (3.7)	19.2 (4.3)	18.4 (3.7)	19.0 (3.7)
Role play (0–20)	305	14.5 (3.5)	14.0 (3.5)	14.7 (3.3)	14.8 (3.3)	391	14.3 (3.4)	14.3 (3.5)	14.1 (3.2)	14.2 (3.3)
Children’s own planning (0–40)	304	23.9 (6.4)	24.3 (5.8)	22.9 (5.8)	24.7 (5.7)	391	22.2 (5.3)	22.8 (6.2)	23.1 (5.2)	23.6 (5.7)
Counting and math activities (0–55)	303	26.5 (9.6)	28.0 (10.3)	26.1 (9.9)	31.0 (8.7)	391	26.3 (11.1)	27.5 (9.9)	27.1 (9.9)	31.4(8.4)
Language activities (0–75)	303	47.4 (13.4)	46.4 (12.4)	45.8 (13.2)	46.8 (12.0)	391	42.7 (13.6)	43.1 (13.6)	43.3 (12.5)	43.6 (12.6)

Notes: The numbers in parentheses (in the left column) are the range of the measure.

able to share knowledge about the implementing of the intervention. The vast majority of teachers (94 % in center-based care and 96 % in family-based care) felt that participating in the intervention gave insights into how to support children’s language and provided additional resources to focus on the interaction with the individual children. Along the same lines, 96 % of teachers in family-based programs and 89 % of teachers in center-based programs indicated that they were inspired to use the intervention’s sequence and scope in their future work. Teachers also felt to a large extent (90 % in center-based childcares and 95 % in family-based childcare) that working with a curriculum gave them insights into their own strengths and weaknesses in providing a rich learning environment to children and that the intervention provided them with new knowledge of how language development can be supported. Interestingly, only 73 % of teachers in center-based program and 80 % of teachers in family-based program responded that they have gained more knowledge about their children’s language development. Finally, we also asked questions addressing the extent to which the participating teachers intended to continue incorporating elements of the “We learn together” intervention in their future educational practice. All teachers in family-based programs and almost all teachers (97 %) in center-based programs were motivated to continue to work with one or more of the elements of “We learn together”.

We conclude that the social validity of “We learn together” was high and that most teachers were inspired to continue to work with core elements of the intervention. There were no significant differences between center-based and family-based in terms of teachers’ attitude to the intervention.

To measure fidelity, we used a measure of child-level intervention exposure (see Bleses et al., 2021 for more details). The recommended quantity of activities for each child each week was participation in a minimum of two large-group activities, two small-group activities and one individual conversation with a teacher. If each child participates in the recommended quantity of activities, we regard the implementation

fidelity as high. Results, as tabulated in Table 5, showed that, on average, teachers in both programs met the expected level of implementing the “We learn together” intervention at the level of the individual child. However, there was some variance in the extent to which participating childcares met the weekly recommendation of at least two small-group activities and one individual child-teacher conversation. Children in family-based childcares participated on average in more small-group activities 2.3 ( $SD = 1.7$ ) compared to children in center-based childcares (1.8 activities,  $SD = 0.9$ ). Even larger differences were found for participation in individual conversations, where individual children experienced an average of 7.3 conversations ( $SD = 5.0$ ) with their teachers each week if they were in a family-based childcare compared to 5.2 conversations ( $SD = 4.0$ ) for children in center-based childcares. In both types of programs, the standard deviations were rather high reflecting large variation at the level of individual children in both types of programs.

4.3. Main effects and moderation by program

The second research question pertained to the main effects of the intervention “We learn together” on children’s language, math and social skill in center-based and family-based childcares and, importantly, the extent to which intervention effects were moderated by type of program.

We estimated 7 models, for the overall sample and for each of the two types of programs, one for each of the child outcomes. The results on the treatment effects of the intervention are shown in Table 6. As can be seen from the top part of the table, the overall effects were significantly positive for productive vocabulary ( $d = 0.10$ ) and receptive vocabulary ( $d = 0.13$ ), math language ( $d = 0.59$ ), numeracy ( $d = 0.34$ ), while no significant difference was found for language use ( $d = 0.03$ ). We found overall significant spill-over effects (effects on a skill not directly targeted in the intervention) on empathy in the model on the overall

**Table 5**  
Exposure of children to activities, mean per week, by type of program.

	Center-based program (N = 980)				Family-based program (N = 594)			
	Mean	SD	10th percentile	90th percentile	Mean	SD	10th percentile	90th percentile
Large group activities	2.7	1.3	1.3	4.3	2.7	1.8	0.8	4.9
Small group activities	1.8	0.9	0.9	2.9	2.3	1.7	0.6	4.2
Individual conversations	5.2	4.0	1.2	11.1	7.3	5.0	2.0	14.0

Notes: The table only includes children in the treatment group, as the exposure for children in the control group per definition is equal to 0. The mean per week is calculated for each child as the total exposure divided by the duration of the intervention period (i.e. 20 weeks). The total exposure is the total number of activities in each category over the full intervention period (for the first period). Since not all children have data on all 20 weeks in the intervention period, the calculated mean per week may be downward biased. If the missing data reflect that the child did not participate or that no activities took place in those weeks, then the calculated mean per week is not downward biased.

**Table 6**  
Estimated treatment effects of the intervention (effect sizes), overall and by type of care.

Outcome	Productive vocabulary	Receptive vocabulary	Language use	Math language	Numeracy	Empathy	Self-reg.
Overall	0.10 (0.03) [0.03, 0.17]	0.13 (0.05) [0.03, 0.24]	0.03 (0.04) [-0.06, 0.11]	0.59 (0.07) [0.45, 0.73]	0.34 (0.07) [0.19, 0.48]	0.10 (0.04) [0.02, 0.17]	0.06 (0.05) [-0.04, 0.17]
	<i>0.004</i>	<i>0.013</i>	<i>0.55</i>	<i>&lt;0.001</i>	<i>&lt;0.001</i>	<i>0.006</i>	<i>0.24</i>
N	2170	1818	2170	2159	2159	2039	2039
Center-based care	0.10 (0.05) [-0.00, 0.20]	0.13 (0.08) [-0.03, 0.28]	0.02 (0.06) [-0.10, 0.15]	0.67 (0.10) [0.47, 0.87]	0.37 (0.11) [0.15, 0.59]	0.13 (0.05) [0.03, 0.22]	0.05 (0.07) [-0.08, 0.18]
	<i>0.05</i>	<i>0.10</i>	<i>0.71</i>	<i>&lt;0.001</i>	<i>0.001</i>	<i>0.008</i>	<i>0.47</i>
N	1353	1073	1353	1338	1338	1255	1255
Family-based care	0.11 (0.05) [0.02, 0.20]	0.13 (0.08) [-0.02, 0.29]	0.02 (0.06) [-0.11, 0.14]	0.50 (0.09) [0.31, 0.68]	0.28 (0.09) [0.10, 0.45]	0.05 (0.05) [-0.05, 0.15]	0.07 (0.07) [0.21]
	<i>0.02</i>	<i>0.10</i>	<i>0.77</i>	<i>&lt;0.001</i>	<i>0.002</i>	<i>0.36</i>	<i>0.36</i>
N	817	745	817	821	821	784	784
p-values for interaction term (overall model)	<i>0.79</i>	<i>0.59</i>	<i>0.95</i>	<i>0.77</i>	<i>0.97</i>	<i>0.34</i>	<i>0.73</i>

Notes: Standard errors in parentheses. 95 % confidence intervals in brackets. *p*-values in italics. Standard errors are adjusted for clustering at center/group level. Missing values are handled by using the ML method (FIML). All estimates are from separate value-added models with covariates and municipality fixed effects included (in the overall model, an indicator for type of care is also included). The effect sizes are in terms of standard deviations of the pre-test for the control group. The final row contains *p*-values for the interaction term between type of program and the intervention indicator added to the overall model.

sample ( $d = 0.10$ ) but not on self-regulation and cooperation. Interestingly, inspecting the results by program, it was revealed that this effect was mainly driven by higher effects for center-based programs only.

We also performed a sensitivity analysis by using multilevel models (HLM) to investigate if it made any difference for the results to take account of clustering at the level of classrooms. This did not lead to any changes to the results reported (at most, the changes in the magnitude of the estimated treatment effects were marginal).

To test whether the effects were different for the two types of programs, we included interaction terms between type of program and the intervention in an additional model for the overall sample. For none of the seven outcomes, we found any significant differences in effects between the two types of programs (with *p*-values ranging from 0.34 to 0.97).

#### 4.4. Effect on educational practices across program

In the third research question, we examined if the intervention had any effect on targeted and non-targeted educational practices in the participating childcares. We estimated separate models for each of the five indices for the teachers' educational activities for the overall sample and for each of the two types of programs. In these models, we controlled for the teachers' age, gender, experience, and education. The results on the treatment effects are shown in Table 7. As shown, there is a significantly positive effect ( $d = 0.32$ ) on the targeted educational

practice, counting and math activities, which is probably reflecting that this educational practice was less common before the intervention. There was no significant effect for the targeted language practices. For the non-targeted educational practices, the effects are positive but not significant.

We then investigated the effect of the intervention on the indices in the two types of programs, and here no significant differences emerged. The effect on counting and math activities was significantly positive for both types of programs.

## 5. Discussion

The current study investigated the extent to which the effects of an infant-toddler school readiness intervention, earlier documented in center-based programs only, generalized to family-based programs, thereby adding to the literature on how to improve quality in different infant-toddler ECE programs. To our knowledge it is an addition to the existing literature that the effects of the same curriculum-based intervention, implemented in both center-based and family-based programs, have been tested for direct comparison. Given that many children around the world are placed in family-based programs, and that only high-quality ECE programs have long-run effects on child outcomes and educational attainment in primary school, the study adds highly valuable knowledge of the generalizability of interventions tested in center-based to family-based programs. In the following, we highlight the main

**Table 7**  
Effects of the intervention on educational practices (effect sizes), overall and by type of care.

Teacher outcome	Play activities	Role play	Children’s own planning	Counting and math activities	Language activities
Overall	0.03 (0.08) [−0.13, 0.20]	0.09 (0.08) [−0.07, 0.24]	0.09 (0.08) [−0.06, 0.24]	0.32 (0.07) [0.18, 0.46]	0.07 (0.07) [−0.06, 0.20]
<i>N</i>	696	696	695	694	694
Center-based care	0.12 (0.11) [−0.10, 0.33]	0.19 (0.12) [−0.04, 0.42]	0.17 (0.10) [−0.02, 0.35]	0.33 (0.11) [0.12, 0.54]	0.12 (0.09) [−0.06, 0.31]
<i>N</i>	305	305	304	303	303
Family-based care	−0.08 (0.13) [−0.33, 0.18]	−0.01 (0.11) [−0.23, 0.20]	−0.01 (0.12) [−0.25, 0.24]	0.33 (0.10) [0.13, 0.52]	0.01 (0.10) [−0.19, 0.21]
<i>N</i>	391	391	391	391	391
<i>p</i> -values for interaction term (overall model)	0.33	0.20	0.21	0.82	0.39

Notes: Standard errors in parentheses. 95 % confidence intervals in brackets. *p*-values in italics. Standard errors are adjusted for clustering at center/group level. All estimates are from separate value-added models with covariates included (teacher characteristics: gender, age, education, and experience; in the overall model, an indicator for type of care is also included). The effect sizes are in terms of standard deviations of the pre-test for the control group. The final row contains *p*-values for the interaction term between type of program and the intervention indicator added to the overall model.

results of the study.

### 5.1. Social validity and implementation fidelity

Implementing systematic interventions is a rather new endeavor for teachers in both center-based and family-based infant-toddler childcares in Denmark. The current study examined how teachers in both center-based and family-based programs experienced their participation in a systematic intervention and how they implemented the intervention. The social validity of the intervention, in particular favorable participant satisfaction, is important as participants need to be on board and supportive of the intervention to ensure high fidelity (Greenwood et al., 2020). The current study demonstrated that the intervention was well received by teachers in both childcare programs and teachers in family-based programs had tentatively more positive attitudes. We speculate that several factors might have contributed to this result. As very few teachers in family-based infant-toddler programs have an educational background in ECE, these teachers may experience a need for a supportive intervention and their readiness for implementing the intervention is higher than for ECE teachers in center-based programs. Moreover, many teachers in family-based programs work alone most of the time and have main responsibility for daily activities. However, implementation of “We learn together” increased collaborative work with colleagues as part of the intervention was intended to be carried out in the family-based childcare groups (groups of teachers who meet in the same regular group). Finally, anecdotal evidence suggests that many teachers in family-based programs were very motivated by being invited to participate in a research project (for most teachers, this was their first time) and contribute to developing evidenced-based knowledge of how best to improve educational quality and, in turn, child outcomes.

The social validity of the intervention may have played an important role for the implementation of the intervention. It is a significant result of the overall study that it was feasible to design an intervention that could be implemented at a satisfactory level in both infant-toddler ECE programs (Bleses et al., 2021). In fact, the secondary analyses conducted as part of the current paper results suggest that the implementation fidelity was higher for teachers in family-based programs for small-group activities and individual conversations. The intervention is designed to flexibly match various local contexts and individual children’s needs as teachers themselves were asked to develop activities of their own choice, which may have supported the implementation flexibility. However, the results indicate that teachers in family-based programs have even more flexibility in organizing small-group activities and individual conversations as they do not need to coordinate with other teachers.

### 5.2. No differential effects on child outcomes across program type

We found that the intervention effects, first established in an early-stage intervention study targeting center-based programs only (Bleses et al., 2020), are generalizable to a new type of ECE program, that is family-based programs. The moderation models showed that the effect of the “We learn together” intervention generally did not differ significantly across program type. The only difference concerned the spill-over effect on empathy in center-based but not family-based programs. This may be related to the fact that scores on social-emotional skills were higher for children in family-based programs already at pretest, and therefore children in center-based care could have had more “headroom” for improving these skills. There are no comparable language interventions studies targeting infant/toddlers in comparable types of care. However, the magnitude of the established effect sizes is slightly higher than those found in a meta-analysis of preschool language and literacy interventions by Chambers and colleagues (2016), which showed average short-term impacts of the magnitude of 0.08 for language.

### 5.3. Educational practices before and after participation in “we learn together”

We obtained self-reported measures of teachers’ educational practices in their usual programs before and after the implementation of “We learn together”. First, we found no large differences on pretest scores between the two types of ECE programs, which is in line with the assessments of process quality in the pilot study, where measures of CLASS toddler demonstrate similar levels of educational practices related to *Engaged Support for Learning*. After having implemented the “We learn together” intervention, posttest scores on educational practices related to counting and math activities were significantly higher than before the intervention; this effect was found for both center-based and family-based programs. On the other hand, scores on language activities were not significantly higher after as compared to before the intervention; this may be related to the somewhat lower effects on language compared to math language and numeracy activities in the “We learn together” intervention itself. Recall that only language and math are directly targeted in “We learn together”. There were no spill-over effects on non-targeted educational practices. We speculate that one potential reason why we do not find effects for language practices is the lack of direct alignment between the content of the intervention and the content of these educational practices. Moreover, each scale mixes items relating to activities and different teacher practices. “We learn language” only offers teachers a sequence and scope and some responsive language strategies to use when communicating with the children and leave the choice of activities up to the teachers. For counting and math, the contents of the “We learn together” intervention and the questionnaire are highly aligned as the items are highly related to math vocabulary that is directed targeted in the intervention and counting activities. It is therefore less of a surprise that there is a relation between

intervention effects and changes in teachers' behaviors as measured by the scale. In relation to language, some items are well-aligned and should be improved by the intervention (such as items for vocabulary instructions) whereas other items are related to singing, rhyming, and other activities. Nevertheless, supporting development of language skills will be an important topic to address in future research.

#### 5.4. Limitations and future directions

The current study has several limitations. First, all but one child outcome measure relied on teacher reports, including a math checklist that was developed as part of this study. Given the size of this study, and the resources that would have been necessary to use for teacher-independent assessments of children, teacher report was identified as the best way to measure child outcomes. Although concerns have been expressed about the validity of providing estimates of children's development based on teacher reports, it is a common methodology in large-scale research, such as the current study, involving the assessment of children's development (e.g., the National Center for Early Development and Learning's Multi-State Study of Pre-Kindergarten, the joint National Center for Early Development and Learning–National Institute for Early Education Research State-Wide Early Education Programs Study, and the Twins Early Development Study). Moreover, research has documented that teacher reports are valid and reliable measures of children's behavioral (Bishop et al., 2003) and academic skills (Justice et al., 2009). While relying on teacher report is a limitation of this study, as it introduces possible biases, evidence suggests that the use of a standardized assessments may result in slightly higher effects; thus our outcomes based on teacher ratings may have underestimated the effects.

Second, long-term follow-up data are needed to determine whether changes in teacher practices are sustained over time. No research of which we are aware has sought to determine if intervention implementation that targets infant/toddler results in lasting change. In addition, long-term follow-up data would also be instrumental in determining whether changes in children's skills may occur following intervention that were not identified using teacher report. Given the limitation of teacher report noted previously, future large-scale studies could follow a randomly selected subset of children for longitudinal standardized assessments to document the potential for positive long-run effects of an infant/toddler intervention.

Finally, even though the impact of the “We learn together” was shown to generalize to a different type of program than originally tested in, it is still unclear whether the intervention effects would successfully generalize to other educational contexts. For instance, it is unclear whether the intervention would be effective in infant-toddler classrooms in international contexts in which teachers in family-based programs have less training and supports.

## 6. Conclusion

The overall aim of the present follow-up study was to explore the effects of an early school readiness intervention—“We learn together”—across center-based and family-based programs. The intervention demonstrated positive impact on children's language, math and social-emotional skills, with the limitation that only one of the measures, due to the age of the children, was based on standardized tests. The analyses show only few differences between the two types of programs. The main contribution of this paper is therefore to demonstrate the feasibility of implementing an evidenced-based school readiness intervention in family-based programs with the same success as in center-based programs. The social validity was high across program and almost all teachers indicated that they would like to continue to implement the intervention after the research project ended. The new insights gained from this study are important as many children worldwide are in family-based programs where studies indicate that quality of the learning environment should be improved to support child

development.

## CRedit authorship contribution statement

**Dorthe Bleses:** Conceptualization, Methodology, Writing – original draft, Project administration, Funding acquisition. **Peter Jensen:** Methodology, Data curation, Formal analysis, Writing – original draft. **Anders Højen:** Conceptualization, Writing – review & editing. **Marinka M. Willemsen:** Conceptualization, Writing – review & editing. **Pauline Slot:** Conceptualization, Writing – review & editing. **Laura M. Justice:** Conceptualization, Writing – review & editing.

## Data availability

Data cannot be shared publicly because part of the data is from administrative registers provided by Statistics Denmark.

## Acknowledgments

This research was supported from a grant by from the Ministry of Children and Education and TrykFonden. We wish to thank the funding bodies, and the participating municipalities, teachers, and children.

## References

- Ackerman, D. (2021). *Reviewing the research base from 1985–2020: Infant and toddler child care and at-risk children's school readiness*. New Brunswick, NJ: Infant and Toddler Policy Research Center at NIEER.
- Alves, J., Guedes, C., & Cadima, J. (2023). The power of teacher-toddler relationships and stability of care for language development. *Social Development*.
- Bauchmüller, R., Görtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly*, 29, 457–470.
- Bigras, N., Bouchard, C., Cantin, G., Brunson, L., Coutu, S., Lemay, L., Tremblay, M., Japel, C., & Charron, A. (2010). A comparative study of structural and process quality in center-based and family-based child care services. *Child and Youth Care Forum*, 39, 129–150.
- Bishop, G., Spence, S. H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*, 74, 1899–1917.
- Bleses, D., Jensen, P., Højen, A., & Dale, P. S. (2018). An educator-administered measure of language development in young children. *Infant Behavior and Development*, 52, 104–113.
- Bleses, D., Jensen, P., Højen, A., Slot, P., & Justice, L. (2021). Implementing toddler interventions at scale: The case of “we learn together”. *Early Childhood Research Quarterly*, 57, 12–26.
- Bleses, D., Jensen, P., Slot, P., & Justice, L. (2020). Low-cost teacher-implemented intervention improves toddlers' language and math skills. *Early Childhood Research Quarterly*, 53, 64–76.
- Blinkoff, E., Nesbitt, K. T., Golinkoff, R. M., & Hirsh-Pasek, K. (2023). Investigating the contributions of active, playful learning to student interest and educational outcomes. *Acta Psychologica*, 238, Article 103983.
- Bratsch-Hines, M. E., Carr, R., Zgourou, E., Vernon-Feagans, L., & Willoughby, M. (2020). Infant and toddler child-care quality and stability in relation to proximal and distal academic and social outcomes. *Child Development*, 91, 1854–1864.
- Bumgarner, E., & Brooks-Gunn, J. (2015). The association between early care arrangements, quality, and emergent bilingual Latino American children's math and literacy skills in English. *Early Childhood Research Quarterly*, 30, 32–44.
- Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12, 3–9.
- Burchinal, M., Garber, K., Foster, T., Bratsch-Hines, M., Franco, X., & Peisner-Feinberg, E. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly*, 55, 35–51.
- Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early childcare and education. *Handbook of Child Psychology and Developmental Science*, 1–45.
- Cates, C. B., Dreyer, B. P., Berkule, S. B., White, L. J., Arevalo, J. A., & Mendelsohn, A. L. (2012). Infant communication and subsequent language development in children from low income families: the role of early cognitive stimulation. *Journal of Developmental and Behavioral Pediatrics*, 33, 577–585.
- Clearinghouse, W. W. (2017). *Standards handbook (version 4.0)*. Washington, DC: Institute of Education Sciences. Retrieved from <https://ies.ed.gov/ncee/wwc/handbooks>.
- Côté, S. M., Mongeau, C., Japel, C., Xu, Q., Séguin, J. R., & Tremblay, R. E. (2013). Child care quality and cognitive development: Trajectories leading to better preacademic skills. *Child Development*, 84, 752–766.
- Chambers, B., Cheung, A. C., & Slavin, R. E. (2016). Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood education: A systematic review. *Educational Research Review*, (18), 88–111.



- Dale, P. S., Paul, A., Rosholm, M., & Bleses, D. (2023). Prediction from early childhood vocabulary to academic achievement at the end of compulsory schooling in Denmark. *International Journal of Behavioral Development*, 47, 123–134.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., & Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428–1446.
- Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *The Journal of Economic Perspectives*, 27, 109–132.
- Burchinal, M., & Farran, D. C. (2020). What does research tell us about ECE programs. *Getting it right: Using implementation research to improve outcomes in early care and education* (pp. 13–36). New York, NY: Foundation for child development.
- Duncan, G., Kallil, A., Mogstad, M., & Rege, M. (2022). Investing in early childhood development in preschool and at home. Working Paper 29985 <http://www.nber.org/papers/w29985>.
- Eckhardt, A. G., & Egert, F. (2020). Predictors for the quality of family child care: A meta-analysis. *Children and Youth Services Review*, 116, Article 105205.
- Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): data from American English and Mexican Spanish infants. *Journal of Child Language*, 35, 77–98.
- Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the computerized comprehension task: Comprehension predicts production. *Developmental Psychology*, 48, 136–148.
- Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). Teaching Math to Young Children. Educator's Practice Guide. What Works Clearinghouse. NCEE 2014-4005. What Works Clearinghouse.
- Greenwood, C. R., Schnitz, A. G., Carta, J. J., Wallisch, A., & Irvin, D. W. (2020). A systematic review of language intervention research with low-income families: A world gap prevention perspective. *Early Childhood Research Quarterly*, 50, 230–245.
- Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, 94, 30–43.
- Gupta, N. D., & Simonsen, M. (2016). Academic performance and type of early childhood care. *Economics of Education Review*, 53, 217–229.
- Hadley, E. B., Barnes, E. M., & Hwang, H. (2023). Purposes, places, and participants: A systematic review of teacher language practices and child oral language outcomes in early childhood classrooms. *Early Education and Development*, 34, 862–884.
- Hadley, E. B., Barnes, E. M., Wiernik, B. M., & Raghavan, M. (2022). A meta-analysis of teacher language practices in early childhood classrooms. *Early Childhood Research Quarterly*, 59, 186–202.
- Harms, T., & Clifford, R. M. (1989). *The family day care rating scale [FDCRS]*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family child care environment rating scale revised edition [FCCERS-R]*. New York, NY: Teacher College Press.
- Helmerhorst, K. O., Riksen-Walraven, J. M. A., Fukink, R. G., Tavecchio, L. W., & Deynoot-Schaub, M. J. G. (2017). Effects of the caregiver interaction profile training on caregiver–child interactions in Dutch child care centers: A randomized controlled trial. *Child & Youth Care Forum*, 46, 413–436.
- Joo, Y. S., Magnuson, K., Duncan, G. J., Schindler, H. S., Yoshikawa, H., & Ziol-Guest, K. M. (2020). What works in early childhood education programs?: A meta-analysis of preschool enhancement programs. *Early Education and Development*, 31, 1–26.
- Justice, L. M., Bowles, R. P., Pence Turnbull, K. L., & Skibbe, L. E. (2009). School readiness among children with varying histories of language difficulties. *Developmental Psychology*, 45, 460–476.
- Koh, S., & Neuman, S. B. (2009). The impact of professional development in family child care: A practice-based approach. *Early Education and Development*, 20, 537–562.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49, 241–253.
- La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom assessment scoring system (CLASS) manual, toddler*. Paul H. Brookes Publishing Company.
- Lally, J. R. (2010). School readiness begins in infancy. *Phi Delta Kappan*, 92, 17–21.
- Landry, S. H., Zucker, T. A., Taylor, H. B., Swank, P. R., Williams, J. M., Assel, M., Crawford, A., Huang, W., Clancy-Menchetti, J., & Lonigan, C. J. (2014). Enhancing early child care quality and learning for toddlers at risk: The responsive early childhood program. *Developmental Psychology*, 50, 526–541.
- Lindeberg, N. H., Hansen, A. T., Vixø, K., Madsen, D. C., Bjerre, L. B., Lundby, A., Mahoney, A. N., & Laurrup, M. A. (2023). *Kvalitet i dagtilbud—National undersøgelse af kvaliteten i pædagogiske læringsmiljøer og rammer i kommunale daginstitutioner og dagplejen for 0-2-årige børn*. VIVE.
- McDaniel, M. E., Townley-Flores, C., Sulik, M. J., & Obradović, J. (2022). Widely used measures of classroom quality are largely unrelated to preschool skill development. *Early Childhood Research Quarterly*, 59, 243–253.
- Merz, E. C., Landry, S. H., Johnson, U. Y., Williams, J. M., & Jung, K. (2016). Effects of a responsiveness-focused intervention in family child care homes on children's executive function. *Early Childhood Research Quarterly*, 34, 128–139.
- Ministry of Children and Social Affairs. (2018a). *Børns alder ved start i dagtilbud*.
- Ministry of Children and Social Affairs. (2018b). *Velfærdspolitisk Analyse: Personale i daginstitutioner - normering og uddannelse*.
- Mitsven, S. G., Perry, L. K., Tao, Y., Elbaum, B. E., Johnson, N. F., & Messinger, D. S. (2022). Objectively measured teacher and preschooler vocalizations: Phonemic diversity is associated with language abilities. *Developmental Science*, 25, e13177.
- Moreno, A. J., Green, S., & Koehn, J. (2015). The effectiveness of coursework and onsite coaching at improving the quality of care in infant–toddler settings. *Early Education and Development*, 26, 66–88.
- Nøhr, K., Dalsgaard, C., Kloppenborg, H., Meldgaard, J., & Bækgaard, M. (2012). *Sådan passer kommunerne børn*. KORA.
- OECD. (2021). *Education at a glance 2021: OECD indicators*. OECD.
- Ofsted. (2020). *Quarterly registration statistics. Childcare providers and inspections as at 31 August 2020* <https://www.gov.uk/government/collections/early-years-and-childcare-statistics>.
- Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement. *Issues in Pre-Kindergarten Programs and Policy*, 19–30.
- Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, 36, 259–268.
- Sjoe, N. M., Bleses, D., Jensen, P., Dybdal, L., Kreiner, S., Sehested, K., & Kirkeby, H. (2019). Measurement properties of the social emotional assessment measure (SEAM) questionnaire using Rasch analysis on data from a representative sample of 0–6-year-olds. *Journal of Psychoeducational Assessment*, 37, 320–337.
- Sjoe, N., Kill, A., Bleses, D., Dybdal, L., Kreiner, S., & Jensen, P. (2020). Assessing strengths and difficulties in social development: a comparison of the social emotional assessment measure (SEAM) with two established developmental psychopathological questionnaires. *European Journal of Developmental Psychology*, 17, 103–122.
- Slot, P. L. (2018). Structural characteristics and process quality in early childhood education and care: A literature review. *OECD Education Working Papers, No. 176*. Paris: OECD Publishing.
- Slot, P. L., Leseman, P. P., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, 33, 64–76.
- Squires, J. (2014). *Social-emotional assessment/evaluation measure (SEAM)*. Paul H. Brookes Publishing.
- Zauche, L. H., Mahoney, A. E. D., Thul, T. A., Zauche, M. S., Weldon, A. B., & Stapel-Wax, J. L. (2017). The power of language nutrition for children's brain development, health, and future academic achievement. *Journal of Pediatric Health Care*, 31, 493–503.
- Vandenbroeck, M., Slot, P., & Hulpia, H. (2021). Quality in home-based childcare providers: variations in process quality. *European Early Childhood Education Research Journal*, 29, 261–277.
- Walker, D., Sepulveda, S. J., Hoff, E., Rowe, M. L., Schwartz, I. S., Dale, P. S., & Levine, S. C. (2020). Language intervention research in early childhood care and education: A systematic survey of the literature. *Early Childhood Research Quarterly*, 50, 68–85.
- Weiland, C., McCormick, M., Matterna, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open*, 4, 1–16.
- Werner, C. D., Vermeer, H. J., Linting, M., & Van IJzendoorn, M. H. (2018). Video-feedback intervention in center-based child care: A randomized controlled trial. *Early Childhood Research Quarterly*, 42, 93–104.
- Von Suchodoletz, A., Lee, D. S., Henry, J., Tamang, S., Premachandra, B., & Yoshikawa, H. (2023). Early childhood education and care quality and associations with child outcomes: A meta-analysis. *PLoS One*, 18(5), Article e0285985.