Early Childhood Research Quarterly 53 (2020) 64–76

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Low-cost teacher-implemented intervention improves toddlers' language and math skills

Dorthe Bleses[a,b,c,*], Peter Jensen[b,c], Pauline Slot[d], Laura Justice[e]

[a] School of Communication and Culture, Aarhus University, 8000 Aarhus C, Denmark
[b] TrygFonden's Centre for Child Research, Aarhus University, 8210 Aarhus C, Denmark
[c] Department of Economics and Business Economics, Aarhus University, 8210 Aarhus C, Denmark
[d] Department of Child, Family, and Education Studies, Utrecht University, The Netherlands
[e] Department of Educational Studies, The Ohio State University, USA

ARTICLE INFO

Article history:
Received 22 April 2019
Received in revised form 27 February 2020
Accepted 1 March 2020

Keywords:
School readiness intervention
Toddler classrooms
RCT study
Language, math, and socio-emotional outcomes
ECE quality

ABSTRACT

The first years of life are foundational for children's cognitive and social–emotional development and to subsequent development in a variety of domains. In many countries, an increasing number of toddlers is enrolled in early childhood education (ECE) programs, which can facilitate early development. However, the quality of these programs is often mediocre. The present study represents a large-scale randomized control trial (87 childcare centers and 1116 toddlers) designed to elevate the quality of instruction in ECE programs serving toddlers. We show that a low-cost 20-week intervention providing teachers with sequence and scope and supportive tools to be more explicit and intentional in their interactions with children resulted in positive, mainly medium- to large-sized effects on targeted language and math skills.

The first years of life are foundational for children's cognitive and social–emotional development. Research on brain development, particularly concerning the plasticity of the developing brain, shows high susceptibility to positive and negative environmental influences during early childhood (Phillips et al., 2017; Shonkoff et al., 2016). This results in substantial variation in children's development at a very early age (Fenson et al., 2007) with profound effects on subsequent development in a variety of domains, which are predictive of academic trajectories through later schooling including early language, math and (to some extent) social–emotional skills – often referred to as "school readiness" (Duncan et al., 2007). Variation in these skills is associated with demographic factors, such as socioeconomic status (SES) and sociocultural background, with children from lower-SES backgrounds and some ethnic minority groups under-performing relative to their more affluent peers (Hoff, 2013).

Given the increased demand for childcare due to the rise of mothers joining the work force (Baker, Gruber, & Milligan, 2008; Burchinal, Magnuson, Powell, & Soliday-Hong, 2015), an increasing number of toddlers are enrolled in early childhood education (ECE) programs in many countries (Organisation for Economic Development & Cooperation [OECD] Family database, 2017). However, research on programs serving the youngest children in the United States (US) suggests mediocre quality (Burchinal et al., 2015). Learning more about how to support learning in such educational settings is therefore of great interest to teachers and policymakers, but research on effective intervention models in toddler classrooms is limited (Burchinal et al., 2015). The present study represents one of few large-scale studies designed to elevate the instructional and process quality of ECE programs serving toddlers and measure its effectiveness on one- and two-year-old toddlers' language and math skills based on teacher ratings.

1. The foundation of school readiness in early childhood

School readiness, which can be conceptualized along domains of pre-academic skills and social–emotional skills, provides the foundation for all subsequent cognitive and social development. Concerning *pre-academic skills*, language in toddlerhood (in particular, vocabulary and the emergence of complex language skills) is an important predictor of later language development (e.g., Fenson et al., 2007) and reading skills (e.g., Bleses, Makransky, Dale, Højen, & Ari, 2016; Dale, Price, Bishop, & Plomin, 2003; Duncan et al., 2007; Lee, 2011). Likewise, early numeracy skills (i.e., knowledge about

* Corresponding author.
E-mail address: bleses@au.dk (D. Bleses).

https://doi.org/10.1016/j.ecresq.2020.03.001

number and quantity) emerge in toddlerhood and lay the foundation for more complex mathematical abilities and problem solving up to six years later (Clements & Sarama, 2011; Duncan et al., 2007). Further, Purpura and Reid (2016) have documented that content-specific language, including words referring to quantities (e.g., "more", "less",), spatial language (e.g., "above" or "beneath") or names of shapes (e.g., "triangle" or "square"), is more important for math development in preschoolers than general vocabulary.

Also, important *social–emotional skills* are being developed in toddlerhood. Denham (2006) identified key aspects of social–emotional competence that are essential for concurrent and future well-being and mental health, which include emotional expressiveness, understanding of emotion, regulation of emotion and behavior, social problem solving, and social and relationship skills. Positive social behavior and emotion understanding in preschool have shown to be related to positive social behavior in kindergarten (Nix, Bierman, Domitrovich, & Gill, 2013).

Importantly, research findings highlight substantial achievement gaps across these areas of school readiness skills that emerge in toddlerhood (see Hoff, 2013). Provision of early childhood programs to young children is viewed as one important avenue for mitigating these achievement gaps.

2. Quality in early childhood education programs

One mechanism for addressing early gaps in school readiness skills is through the provision of high-quality ECE programs. There are important structural differences between ECE settings, which serve infants and toddlers, and preschool settings serving 3- to 5-year-olds. For instance, the latter has a higher teacher–child ratio and likely a greater emphasis on instructional routines and skills via structured curricula. However, both settings can be viewed through a lens of socioecological and transactional frameworks, which incorporate multidimensional and bidirectional influences of both children and their environment in shaping developmental processes (Burchinal et al., 2015). Meta-analytic evidence finds overall positive impacts of early ECE program enrollment on children's language, literacy and math skills in the early years (Duncan & Magnuson, 2013), in some cases also longitudinally (Gupta & Simonsen, 2010; Vandell, Burchinal, & Pierce, 2016). For instance, Vandell et al. (2016) found that more experience in center-based care was associated with higher academic grades and admission to more selective colleges, although lack of persistence and fadeout in impacts is a concern in such work (Bailey, Duncan, Odgers, & Yu, 2017; Phillips et al., 2017). However, the quality of the ECE settings is crucial (Burchinal et al., 2015).

Based on more specific developmental theories of cognitive and social stimulation and attachment theory, several hallmarks of high quality ECE have been forwarded as important for promoting cognitive and social skills and both structural and process quality have been shown to be related to child outcomes (Burchinal et al., 2015). Structural quality is defined as the regulable, more distal aspects of ECE, such as group size, teacher-child ratio, or teachers' qualifications, which are regarded as important preconditions of process quality, whereas process quality concerns children's daily experiences and interactions with teachers and peers are proximal determinants of children's outcomes (Burchinal et al., 2015). In particular high-quality interactions between child-teacher and rich learning opportunities have been shown to be the most crucial (Burchinal et al., 2015; Slot, 2018). Medium- to high-quality teacher–child interactions are, for instance, associated with higher social–emotional and cognitive gains in preschool-aged children (Hatfield, Burchinal, Pianta, & Sideris, 2016). The quality of teacher–child interactions in ECE programs varies extensively across programs both in the US and in Europe (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Slot, 2018).

However, most of the work on quality referred to thus far, is based on research conducted in preschool settings. The developmental trajectories of preschool children are rooted in infant-toddlerhood experiences, and in particular, social–emotional and language skills emerge and develop quickly during infancy and toddlerhood. Accordingly, the type of teacher–child interactions in infant/toddler classrooms that have received most attention is building relationships, fostering cognitive development and encouraging language development (Norris & Horm, 2015). Support for social–emotional development, including self-regulation and behavior guidance, may therefore be even more important in this age range compared to the preschool years (Thomason & La Paro, 2009). This emphasis on such "soft skills" as social–emotional development and interpersonal and personal skills also reflects parents' view of which developmental areas toddler settings should focus on, whereas for preschool aged children more emphasis is put on preacademic skills, as highlighted in a cross-European study among about 2000 parents (Broekhuizen, Leseman, Moser, & van Trijp, 2015). High structural quality is rated equally important in infant/toddler and preschool classrooms, with having a safe environment being rated as the most important (Broekhuizen et al., 2015). In addition, parents rated process quality as slightly more important than the structural features (Broekhuizen et al., 2015). In US infant/toddler classrooms, the quality of emotional and behavioral support and support for learning has been reported to be lower than in preschool classrooms. with studies reporting minimal levels of quality for infant/toddler (Burchinal et al., 2015; King, Pierro, Li, Porterfield, & Rucker, 2016; La Paro, Williamson, & Hatfield, 2014); however, it is important to note that relatively little is known about the effects of center-based ECE programs on infants and toddlers (Burchinal et al., 2015).

3. Improving quality and child outcomes in toddler classrooms

Since many ECE programs provide only mediocre quality of adult–child interactions, particularly concerning instructional support, numerous intervention studies have sought to elevate the nature of these interactions. These often involve providing teachers with advanced training and curricula that help them provide higher-quality and more intentional interactions with children (Markussen-Brown et al., 2017). Many preschool-focused interventions are multi-component curricula, which specify a scope and sequence delineating a set of specific targeted skills organized over the period of instruction, and instructional practices (i.e., techniques and aligned activities), often manualized via scripted lessons that teachers implement with the children (Justice et al., 2010; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011). However, a recent study has indicated that the most essential component in a language and emerging literacy intervention implemented at scale to improve process quality was an explicit sequence and scope of learning objectives, whereas the provision of scripted lessons was less important (Bleses, Højen, Dale et al., 2018).

The majority of this work focuses on teachers working with preschool-aged children, with work concentrated on settings serving infants and toddlers being rare. Some exceptions include studies that have tested the effectiveness of various professional development interventions (of coursework, onsite coaching or video-based feedback) aimed at improving the quality of interactions in center- and home-based care (e.g., Helmerhorst, Riksen-Walraven, Fukkink, Tavecchio, & Deynoot-Schaub, 2017; Moreno, Green, & Koehn, 2015; Werner, Vermeer, Linting, & Van IJzendoorn, 2018). For instance, Moreno and colleagues showed that a combination of coursework and coaching demonstrated patterns of improvements in teacher practices related to the quality of interactions. These

studies suggest that the quality of interactions can be improved, but none of these studies measured the effects of improved quality on child outcomes (Helmerhorst et al., 2017; Moreno et al., 2015; Werner et al., 2018). Moreover, none of these studies tested a specific curriculum incorporating age-appropriate sequence and scope or aligned stimulating activities.

One study did evaluate the effect of a curriculum (Responsive Early Childhood Curriculum, RECC), that supported quality (by encouraging teachers to respond contingently to children's signals) and age-appropriate, stimulating activities to promote language, early literacy, and math development in toddlers (Landry et al., 2014). The RECC thematic curriculum included four daily cognitive activities led by the teacher in a large group (two read alouds and nursery rhyme) as well as one small group activity (math) implemented over a 36 week-period (Landry et al., 2014). An additional social–emotional curriculum, tested in a separate arm, was based on an explicit sequence and scope. The study focused on high-risk children. Teachers in the intervention arms outperformed control teachers on measures of rich language input, classroom organization and planning, and implementation of engaging learning activities. No significant group differences were found for growth in children's pre-academic skills, but children in the intervention arms outperformed controls in areas of social and emotional development, indicating that sequence and scope might be an important part of promoting development for this age range (Landry et al., 2014). In summary, only a few studies have examined how the quality of infant and toddler classrooms can be improved and even fewer studies have documented effects on child outcomes.

4. Differential effects for different groups of children

A general limitation in the intervention literature (independent of age range) is that the study sample sizes are small (10 children or fewer) and/or many studies focus entirely on children from low-SES homes (Walker et al., 2020); both factors prevent examination of potential differential effects conditional on child and parental background characteristics. To the best of our knowledge, no study has investigated differential effects of the same toddler school readiness intervention on children from varying backgrounds. Earlier work from the implementation of the same language and literacy interventions targeting preschool aged children in Denmark suggest no differential effects between children depending on child and family characteristics (Bleses, Højen, Dale, et al., 2018; Bleses, Højen, Justice, et al., 2018), but it is unclear to what extent this finding also applies to toddlers who are in earlier stages of development.

5. ECE settings in Denmark

Compared to the US childcare/preschool system, the Danish childcare system features nearly universal enrollment of children and all municipalities are obliged to ensure availability of childcare (Gupta & Simonsen, 2010). According to the OECD family database, Denmark has the highest percentage of children enrolled in infant and toddler care among OECD countries (OECD Family Database, 2017). Such extensive investment in ECE programs by the local and federal government is consistent with the design of the Danish welfare state, where nearly all women and men are on the labor market. Three out of four children are enrolled in childcare before they turn three years and are, on average, about ten months old when they start in childcare (Ministry of Children and Social Affairs, 2018a). The majority of children are enrolled in center-based care (63%), and most serve children in the entire age range from birth to five years (Justice, Logan, Purtell, Bleses, & Højen, 2017); often, infant-toddler classrooms are separated from preschool classrooms with typically eight to 12 children (Nøhr, Dalsgaard, Kloppenborg, Meldgaard, & Bækgaard, 2012). The socioeconomic composition of childcare centers is mainly determined by the composition of the neighborhood. Childcare centers vary greatly in size (range 11–80 children), but most enroll an average of 40–50 children (Nøhr et al., 2012). The public spending on childcare in Denmark is the 5th highest among OECD countries (OECD Family database, 2017) and the structural quality is regarded high (Slot, 2018). For childcare centers serving infants and toddlers, the teacher–child ratio is relatively low, with one teacher to about 3.5 children and approximately 60% of teachers have a 3.5-year pedagogical bachelor degree, but with some local variation (Ministry of Children and Social Affairs, 2018b).

Typically, the educational practice in Danish childcare centers is based on a child-oriented whole-child perspective and teachers are trained to take a holistic approach to pedagogy (Bauchmüller, Gørtz, & Rasmussen, 2014). Encouraging social skills is the hallmark of the Danish educational practice, whereas an explicit focus on domain-specific academic skills is uncommon.

A recent review, which examined the relation between structural and process quality in preschools in OECD countries using the *Classroom Assessment Scoring System PreK* (CLASS; Pianta, La Paro, & Hamre, 2008), placed Denmark in the high range for Emotional Support, in the medium range for Classroom Organization, whereas scores for Instructional Support were in the low range, which is in line with international evidence (Slot, 2018; Slot, Bleses, Justice, Markussen-Brown, & Højen, 2018). In summary, 75% of toddlers are enrolled in ECE programs with relatively high structural quality, but this does not necessarily translate into high process quality.

6. The present study

The present study represents one of the only large-scale randomized control trial (RCT) designed to elevate the quality of instruction in ECE programs serving toddlers to increase the intentional support of toddlers' school readiness. The intervention targeted toddlers' language and math skills, and measures effects on these domains. As social–emotional skills and other "soft skills" are already well supported in the Danish context (see Slot et al., 2018), we elected to not focus on promoting this domain directly in the intervention and rather to focus on improving the low language and cognitive stimulation of the settings. However, potential spillover effects of the intervention are also measured on social–emotional skills. Next, we present several key characteristics of the study.

First, the study was designed to be a low-cost teacher-implemented intervention "Play and Learn" to be used in the context of universal childcare in Denmark without providing any additional resources to the centers. The effect of the intervention is tested in an unselected heterogenous sample of 1116 toddlers from five municipalities, which allows us to investigate not only main effects on child outcomes, but also differential effects on subgroups of children. Dual language learners (DLLs) in Denmark (children of immigrant parents in Denmark, for who Danish is a second or later language) tend to originate from Turkey, Lebanon, Pakistan, Iraq, and Somalia (Statistics Denmark, 2014). Concerns about achievement gaps of socially disadvantaged children and DLLs are present in the Danish context, and research indicates that by the time children leave infant-toddler classrooms, the 15% of children with weakest language skills are one year behind the 15% children with the strongest language skills, a gap that is doubled by the end of preschool (Bleses, Jensen, Nielsen, Sehested, & Sjoe, 2016).

Second, the intervention targets both teachers' instructional and interactive skills and children's language and math skills; thus, the theory of change is such that the intervention aims to increase intentional teaching, which, in turn, would improve children's skills. Research finds that effective preschool curricula are typically multi-component, comprising a scope and sequence delineating a

set of specific targeted skills organized over a period of instruction, and aligned instructional practices. To promote fidelity to the components, curricula are manualized via semi-scripted lessons that teachers implement with the children (Chambers, Cheung, & Slavin, 2016). Research also finds that "serve and return" interactive practices, such as contingently responding to children's comments and engaging children in extended conversations are very important predictors of child improvement (Phillips et al., 2017). Importantly, direct training of teachers may increase their use of these practices (Markussen-Brown et al., 2017). Thus, in the design of this intervention, we decided to emphasize the use of a sequence and scope to organize classroom learning targets and teachers' use of responsive strategies.

Accordingly, the "Play and Learn" intervention has a focus on a set of interrelated school readiness skills with training and materials targeted at the teachers, designed to increase high quality instruction and interactions. Borrowing from the most successful preschool interventions, the intervention has a significant focus on specific instructional content using a sequence and scope of selected developmental school readiness targets. However, recent research finds that teacher-implemented curricula effects are heightened if teachers are provided with some discretion in implementation (Bleses, Højen, Dale, et al., 2018). Hence, teachers were trained in using responsive strategies to increase process quality, but were given a high discretion in terms of developing and implementing the educational activities, that is, the sequence and scope was not supplemented with scripted lesson plans with aligned activities. Instead, a threefold scope of instruction targeting language, math language and numeracy was sequenced over a 20-week curriculum period, and teachers were asked to address these skills in self-developed activities of their own choice, and in child-initiated activities and during daily classroom routines.

Three research questions were addressed: (1) To what extent does a cost-efficient intervention improve children's outcomes in language, math language, and math? (2) To what extent are the impacts of "Play and Learn" conditional on child- and parental background characteristics, namely language background and family SES; and (3) To what degree are intervention impacts influenced by the quality of the childcare prior to the intervention?

Table 1
Characteristics of the included childcare centers, classrooms, teachers, children and parents.

Variable	Control	Intervention	Balancing tests
	Total	Total	
Childcare characteristics			
Childcare, *N*	44	43	
Classrooms, *N*	86	90	
Number of children, *N*	578	538	
	Mean	Mean	*p*-values
Teacher characteristics			
Age (years)	40.2	40.4	.92
Experience (years)	10.9	11.4	.65
Education BA-level (fraction)	0.93	0.91	.65
Missing data (0/1)	0.07	0.04	.45
Parent characteristics			
Mother low education (0/1)	0.11	0.13	.49
Mother low/mid education (0/1)	0.28	0.28	.96
Mother high-mid education (0/1)	0.30	0.29	.67
Mother high education (0/1)	0.24	0.23	.83
Mother missing data on education (0/1)	0.07	0.08	.85
Father low education (0/1)	0.17	0.14	.28
Father low/mid education (0/1)	0.35	0.37	.64
Father high-mid education (0/1)	0.19	0.20	.71
Father high education (0/1)	0.21	0.22	.83
Father missing data on education (0/1)	0.08	0.08	.88
No employment (income base) (0/1)	0.13	0.11	.31
Parent cohabiting (0/1)	0.29	0.31	.50
Parents married (0/1)	0.56	0.55	.85
Living with single parent (0/1)	0.15	0.13	.47
Child characteristics			
Boy (0/1)	0.51	0.52	.67
Non-Western origin (0/1)	0.09	0.12	.31
Age (months)	24.9	24.0	.02*

Notes: Education levels: low (primary school); low–mid (high school, vocational education); high–mid (e.g., professional BA such as teacher); high (BA and advanced university education). Balancing tests are for equality of means in the control and intervention group. The *p*-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the childcare center level.

* $p < 0.05$.

7. Method

7.1. Design of the study

This multisite, RCT was designed using a priori power calculations drawing upon estimates from previous studies within education research (Bleses, Højen, Justice, et al., 2018), showing that for the cluster-randomized trial, we would need 82 childcare centers to be able to detect an effect size of 0.2 SDs with a power of 80% and a standard significance level of 5%. However, the considerable variation in size (number of children enrolled) of the childcare centers made it desirable to recruit more childcare centers for the trial to retain the desired power. The classrooms of the included childcare centers had between four and 21 children enrolled, with the 10–90 percentile range being 10–16 children. In the current study, we only include children aged 18 months or older. Thus, 1508 children were enrolled in the study and completed pretest assessments. Data were collected between September 2015 and June 2016.

The RCT design was approved by the five municipalities. The Danish Data Protection Agency approved the collection and treatment of all data for the project (approval no. j.nr. 2014-54-0822). All teachers and parents were informed about the RCT and data collection activities, with materials provided in five different languages, and were assured that all data would be treated anonymously and confidentially. Moreover, parents, professionals and associated childcare centers were informed about the main results to honor their contribution to the work.

7.2. Attrition following random assignment

The RCT used a cluster randomized design with random assignment at the level of the childcare center; each center was randomly assigned to a treatment or a control group, with stratification based on the percentage of socially disadvantaged children in each center. Table 1 shows baseline characteristics of teachers and children and the balance between treatment and control groups for the final sample. As expected in large-scale trials, this study had some missing data. During the intervention period, 11% of these children left their childcare centers, typically because of transition to preschool. An additional 15% of children did not complete posttest assessment. Thus, the attrition rates were 30% (treatment group) and 22% (control group), however, these attrition rates are not significantly different when we adjust for clustering at the childcare center level ($p = 0.091$), so no differential attrition could be detected. Fig. 1 provides a flow-chart of how teachers, classrooms, and children were recruited and enrolled into the study. The final sample involved 87 childcare centers serving 1116 toddlers, in the age range from 18 to 36 months at the time of the pretest assessment. By using the personal identification numbers of the Danish civil registration system, we were able to obtain information from administrative registers on children and family background through Statistics Denmark (see

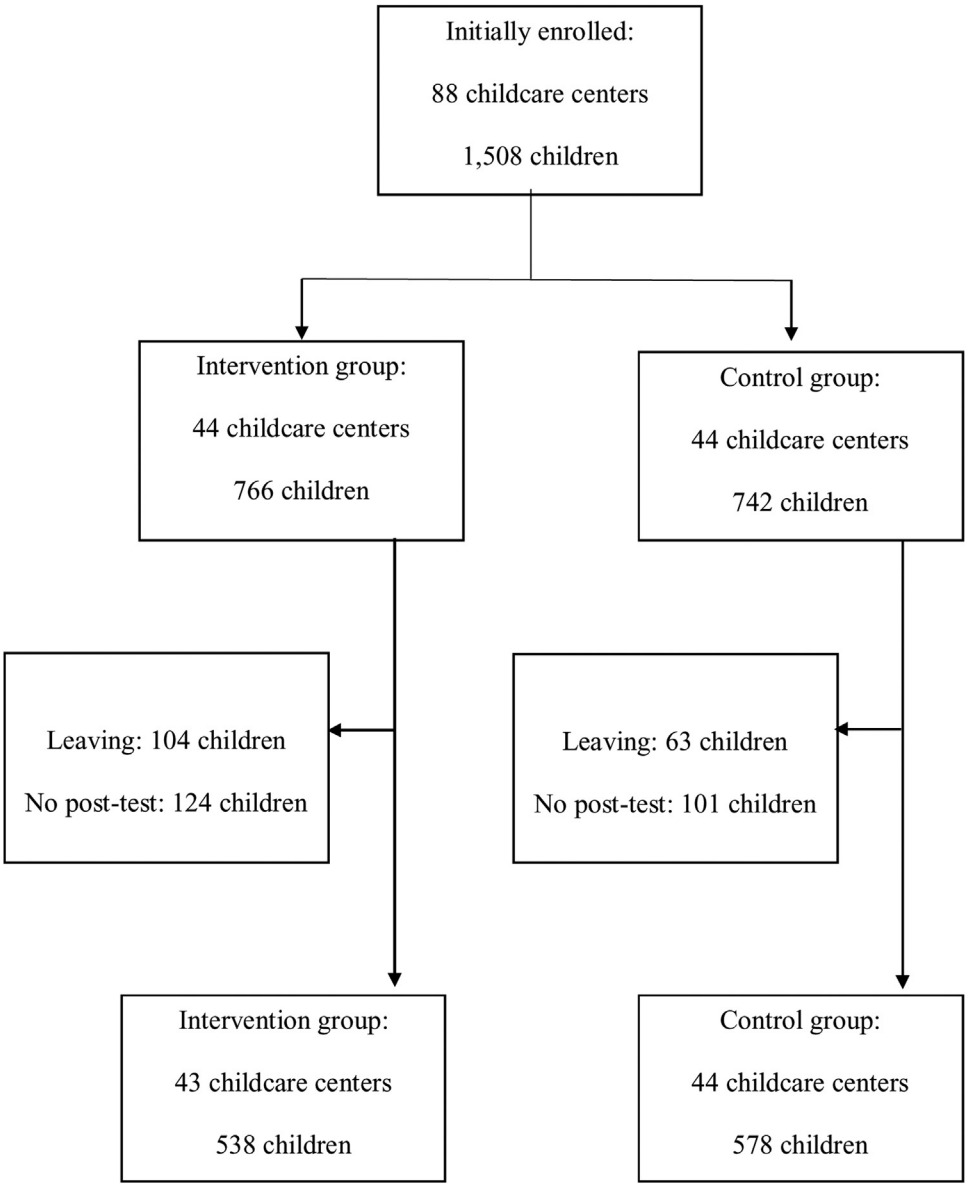

Fig. 1. Flow chart of the recruitment and follow-up/analysis process.

Table 1). Balancing tests demonstrated that the baseline characteristics of the included teacher, parents and child characteristics were distributed evenly across the intervention and control groups for the final sample. The main research questions were evaluated for all children who had a pretest and a posttest on a given subscale (see Table 2). In general, the children in the control group had higher pretest values than the children in the intervention group, with part of the difference being due to the difference in age. This lack of balance is taken care of by modeling the changes in the outcome measures and including age as a covariate (see further below in the Analytic Strategy).

We compared attriters with non-attriters by balancing test between the two groups, both for the total sample and separately for control and intervention groups. The results of these tests are reported in Table A1 and demonstrate that the demographic characteristics do not differ between attriters and non-attriters. We also performed balancing tests for the allocation to the control and interventions groups for the initial sample (i.e., before attrition) and these are reported in Table A2. They demonstrate that the characteristics were also evenly distributed in the initial sample.

7.3. Procedures

The primary procedures of relevance were intervention implementation and implementation of main study measures.

Intervention. All childcare centers in the treatment group implemented a researcher-developed 20-week instructional curriculum called "Play and Learn," the design of which was influenced by recent findings within the field of ECE. The conceptual model behind the intervention is depicted in Table 3. In the following the intervention targets, intervention elements and expected outcomes are described in more detail. The intervention was developed in close partnership with the involved municipalities.

The intervention targets both teacher and child skills. For teachers, the intervention aimed at improving their instructional and interactive skills. The aim of the intervention for children was to improve their language and math skills. Based on theory as well as empirical studies, the intervention had several elements, which were developed to support teachers' and children's skill development. First, we developed a threefold scope of instruction targeting language (general vocabulary and language use) and math (math

Table 2
Descriptive information on children's language, math, and social–emotional skills at pretest and posttest.

	Mean	SD	Range	*N*
Pretest				
Control				
Vocabulary (*CDI-Educator*)	31.3	20.8	0–70	578
Language use (*CDI-Educator*)	4.5	3.2	0–10	578
Math language (*Math-checklist*)	43.0	14.6	28–112	575
Numeracy (*Math-checklist*)	19.0	7.7	13–52	575
Empathy (*SEAM toddler*)	13.0	4.2	0–18	572
Self-regulation, cooperation (*SEAM toddler*)	8.5	2.7	0–12	572
Intervention				
Vocabulary (*CDI-Educator*)	24.7	17.1	0–70	538
Language use (*CDI-Educator*)	3.8	2.8	0–10	538
Math language (*Math-checklist*)	37.5	9.5	28–112	538
Numeracy (*Math-checklist*)	17.0	5.3	13–52	538
Empathy (*SEAM toddler*)	12.3	4.3	0–18	538
Self-regulation, cooperation (*SEAM toddler*)	8.3	2.7	0–12	538
Posttest				
Control				
Vocabulary (*CDI-Educator*)	48.4	17.7	0–70	578
Language use (*CDI-Educator*)	7.2	2.8	0–10	578
Math language (*Math-checklist*)	57.7	19.0	28–112	575
Numeracy (*Math-checklist*)	26.0	9.5	13–52	575
Empathy (*SEAM toddler*)	15.5	3.1	0–18	438
Self-regulation, cooperation (*SEAM toddler*)	9.8	2.2	0–12	438
Intervention				
Vocabulary (*CDI-Educator*)	47.4	17.5	0–70	538
Language use (*CDI-Educator*)	7.1	3.0	0–10	538
Math language (*Math-checklist*)	63.8	19.6	28–112	538
Numeracy (*Math-checklist*)	28.2	9.9	13–52	538
Empathy (*SEAM toddler*)	15.7	3.1	0–18	465
Self-regulation, cooperation (*SEAM toddler*)	9.7	2.5	0–12	465

language and numeracy skills) that was sequenced over the 20-week curriculum period with the overall aim of promoting the instructional quality and to promote teachers' use of rich language. Teachers were asked to develop activities of their own choice to address the instructional targets to promote active engagement and to support autonomy. Examples of the sequence and scope, which has been developed based on empirical databases (e.g., Jørgensen, Dale, Bleses, & Fenson, 2010) include (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 language*: words for numbers, shapes (e.g., "round", "square"), sizes (e.g., "big", "short"), and patterns (e.g., "dots", "stripes"). Pictures and posters of target words were provided to assist teachers in their planning of activities but it was voluntary to use them.

Second, to support teachers' development of intervention activities and the use of rich language in both planned activities and during daily routines, the intervention was organized in five content themes ("My daily life"; "The weather"; "Animals and nature"; "My family"; "Out in the world"). To guide implementation, in particular to support the self-organization of activities and the inclusion of all children, teachers received a weekly planner offering an instructional framework by which to plan these activities in alignment with the week objectives; the planner provided a suggested weekly dosage of two large-group and two small-group activities per week and one individual conversation per child each week (among others to minimize effects of implicit biases in engaging children in activities). The teachers were encouraged to use daily routines like diaper change and dressing situations for individual conversations with children. Child participation in all three types of activity settings was logged on the weekly planner.

Third, to support high interaction quality and use of responsive strategies in all activities during the day (i.e., planned activities as well as daily routines), teachers were provided with a set of strategies to enrich and differentiate their interactions with individual children. The strategies were visualized on posters to be put on the wall. One poster ("The high-quality conversation") included strategies like repetition and expansion, and open-ended questions. Inspired by, for instance, Beck and McKeown (2007), the second poster ("Learning new words") included word learning strategies like relating new words to known words, and providing examples from the same or related semantic categories. The third poster ("The learner's ladder") included examples of targeted scaffolding (low and high support strategies cf., Justice et al., 2010). Both access to posters and videos supported the use of specific practices supporting enriched conversations.

Fourth, based on research indicating the importance of self-evaluation and reflection when implementing instructional practices (Crawford, Zucker, Van Horne, & Landry, 2017), teachers in each classroom completed implementation notes on a logging tool and monitored child gains from the intervention to support reflection on implementation. The implementation logs represented the experiences of individual children in each classroom.

Finally, inspired by recent work on professional development (Hamre et al., 2012), staff in all classrooms participated in a two-day training workshop, which introduced teachers to the intervention elements, including the instructional scope and sequence, the use of specific practices to enrich and differentially support children, and guidance on how best to select activities that create many learning

Table 3
Conceptual model for "Play and Learn".

Intervention targets	Intervention elements	Excepted outcomes
TEACHERS→ Instructional and interactive skills.	Age dependent sequence and scope of → Weekly language, math language and numeracy targets support teachers' instructional quality and use of content rich language. Five content themes support content rich language about events relevant to children's lives. A weekly planner supports the organization of activities and the inclusion of all children. Accessible posters that support high quality interactions (strategies like repetition and expansion, open-ended questions, explanation of new words) and differentiated instruction via targeted scaffolding (low and high support strategies cf. Justice et al., 2010). Weekly implementation notes and monthly child progress checklists that support reflection over implementation and child benefits. Access to resources (materials, videos).	TEACHERS More instructional content, use of more rich language and use of more instructional and interactive strategies. ↓
CHILDREN Child language and math skills.		CHILDREN Improved outcomes in children's language, math language and numeracy.

opportunities. This training provided teachers with all materials necessary to fully implement the 20-week "Play and Learn" curriculum.

Teachers' ongoing completion of implementation logs allowed fidelity to be monitored and this indicated that teachers implemented the "Play and Learn" activities with varying degrees of fidelity. For the weekly large-group activities, of which two were recommended, teachers implemented an average of 1.8 (range 0–5.8). For the weekly recommendation of two small-group activities, teachers implemented an average of 1.1 (range 0–3.9). Finally, individual children experienced an average of about 1 (0.9) individual conversation with their teachers each week (range 0–11.5). Additional measures of fidelity were coded from two videos collected in each classroom mid-way through the implementation period. Analyses showed that treatment teachers used targeted strategies at significantly higher rates than controls (strategies from "The high-quality conversation" were used 4.7 times across teachers in the treatment group vs. 1.0 time in the control group, strategies from "Learning new words" were used on average 1.2 times in the treatment group vs. 0.2 times in the control group and strategies from "The Learner's ladder" were on average used 4.2 times in the treatment group vs. 0.2 times in the control group).

7.4. Measures

Classroom measures. To examine the potential interaction between treatment effects and baseline center process quality, the *Classroom Assessment Scoring System-Toddler* (CLASS-Toddler; La Paro, Hamre, & Pianta, 2012) was used. CLASS is an observational instrument that evaluates the quality of classroom processes along two broad domains, which comprises five and three dimensions, respectively: Emotional and Behavioral Support (EBS, comprising Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Student Perspectives, and Behavior Guidance) and Engaged Support for Learning (ESL, comprising Facilitation of Learning and Development, Quality of Feedback, and Language Modeling). Classroom quality was rated on a 7-point scale ranging from 1 or 2 (classroom is low on that dimension); to 3, 4 or 5 (classroom is in the midrange on that dimension); and to 6 or 7 (classroom is high on that dimension). All observers were trained by a licensed CLASS trainer and achieved at least 80% agreement within one scale-point deviation with the trainer on an online test before they were admitted to assess the classroom quality. A total of 13% of classrooms were randomly selected and assessed by two observers, revealing high inter-rater reliability of 97% within one-scale point (range 87–100%).

If there were only one or two classrooms at the center, we administered CLASS in all classrooms in the center ($n = 88$). In childcare centers with more than two classrooms ($n = 27$), the classrooms to be observed were chosen randomly. Overall, the quality of 97 classrooms out of 176 classrooms were observed using CLASS-Toddler prior to the introduction of the intervention. There were no significant differences on teacher characteristics (i.e., teacher age, educational background and experience) between classrooms that were observed and those that were not. There was one minor difference between the classrooms in terms of child characteristics, such that there were more children from unemployed families in classrooms observed with CLASS (14% vs. 9%, $p = .03$).

Child outcomes. To measure child outcomes, we used three teacher-reported instruments that were completed at pretest (before the intervention) and posttest, approximately seven months after the intervention was initiated. Each of the instruments took on average about eight to 10 min to complete, thus teachers spent about 30 min per child completing the instruments.

For *language*, we used a teacher-based standardized checklist, the *CDI-Educator* (Bleses, Jensen, Højen, & Dale, 2018), developed on the basis of the well-known *MacArthur-Bates Communicative Development Inventories* (Fenson et al., 2007) and a Danish adaptation of a short version of the instrument (Vach, Bleses, & Jørgensen, 2010). The checklist includes a 70-item vocabulary checklist, which is divided in nine categories of content (sound effects and animal sounds, animals and things, food and drink, body parts, small household items, furniture and rooms and places (to go), people and routines, action words, descriptive words, and particles). Moreover, five questions concerning the child's use of decontextualized language with respect to objects and actions distant from the here and now (e.g., whether the child at any time speaks about earlier episodes and persons who are not present or about something that will happen in the future) are included. The vocabulary summary score was calculated adding up the number of words the child could produce. For the language use summary score, the response categories were converted to points and summed up across the five questions ("not yet" [0], "sometimes" [1] or "often" [2]). Measurement properties of the instrument are reported in a validation study (Bleses, Jensen, et al., 2018).

For *math language* and *math*, we used a researcher-developed teacher-administered checklist reflecting the fact that we were not able to identify a standardized measure that could be applied to toddlers. The constructs included in the checklist were primarily based on more general work on children's math development (Frye et al., 2013) and more specific work on content-specific language by Purpura and Reid (2016) that is supportive of math development. The checklist consists of 41 items in total, where 13 items assess children's early numbers and counting skills and 28 items assess children's comprehension and use of math language (words for sizes, quantities, shapes, and space). The response categories were converted to points and summed up across the relevant questions ("not yet" [1], "sometimes" [2], "often" [3] or "always" [4]). This instrument covered items that were more proximal to the intervention targets. Internal consistency was assessed for each subtest. For the subtests, all items have an item-total correlation exceeding .50, and as these individual item correlations suggest, Cronbach's alpha is high for both subtests, around .95. This demonstrates that the math checklist has high internal consistency. Correlations with the two CDI-based language scores (vocabulary, language use) are also substantial, between .60 and .73.

For measuring potential spill-over effects on children's *social–emotional skills*, we used the teacher version of a Danish adaptation (Sjoe et al., 2017) of the standardized questionnaire, *Social–Emotional Assessment/Evaluation Measure (SEAM)–Research Edition* (Squires, 2014). SEAM includes 10 domains (called benchmarks) critical to social–emotional skills and executive functions: empathy, healthy interactions, expression of emotions, regulation of social–emotional responses, cooperation, sharing and engaging, regulation of attention and activity level, independence, self-image, and adaptive skills. SEAM has four response categories: 0 = not true, 1 = rarely true, 2 = somewhat true, and 3 = very true. A high score indicates a positive aspect of social–emotional development. The ten benchmarks were combined into two overall indexes: (1) The empathy index, which assesses the child's ability to communicate own feelings and to read and understand others' feelings; and (2) The self-regulation & cooperation Index, which assesses the child's ability to regulate and cooperate, and the child's adaptability.

7.5. Analytic strategy

To estimate the effects of the intervention compared with treatment as usual in the control group, we use regression analysis with SEs adjusted for clustering at the childcare center level to account

Table 4
Treatment effects (effect sizes) for all children and subgroups.

	Vocabulary	Language use	Math language	Numeracy	Empathy	Self-regulation & cooperation
All children	0.27**	0.19**	0.80**	0.55**	0.12	−0.00
(*n* = 1080)	(0.07)	(0.07)	(0.13)	(0.12)	(0.07)	(0.10)
Subgroups						
Background						
Danish	0.29**	0.23**	0.88**	0.61**	0.14	0.01
(*n* = 972)	(0.08)	(0.08)	(0.14)	(0.13)	(0.07)	(0.10)
Non-Western	0.15	−0.17	−0.01	−0.05	−0.13	−0.17
(*n* = 108)	(0.19)	(0.20)	(0.23)	(0.27)	(0.24)	(0.25)
	p = 0.535	*p = 0.067*	*p = 0.001*	*p = 0.025*	*p = 0.255*	*p = 0.477*
Maternal Ed[a]						
Low	−0.06	−0.23	0.56*	0.49*	−0.23	−0.21
(*n* = 130)	(0.13)	(0.16)	(0.24)	(0.23)	(0.21)	(0.26)
Low–mid	0.37**	0.23*	0.79**	0.53**	0.04	−0.07
(*n* = 307)	(0.12)	(0.12)	(0.14)	(0.13)	(0.14)	(0.14)
High-mid	0.23*	0.16	0.69**	0.44*	0.21*	0.07
(*n* = 319)	(0.10)	(0.12)	(0.16)	(0.18)	(0.10)	(0.13)
Vs. Low	*p = 0.062*	*p = 0.042*	*p = 0.596*	*p = 0.834*	*p = 0.046*	*p = 0.332*
High	0.35**	0.39**	1.15**	0.77**	0.15	0.03
(*n* = 261)	(0.11)	(0.11)	(0.25)	(0.23)	(0.13)	(0.16)
Vs. Low	*p = 0.014*	*p = 0.001*	*p = 0.044*	*p = 0.334*	*p = 0.114*	*p = 0.433*
Vs. High–mid	*p = 0.404*	*p = 0.196*	*p = 0.065*	*p = 0.211*	*p = 0.700*	*p = 0.804*

Notes. Standard errors in parentheses. Standard errors are adjusted for clustering at the childcare center level. Missing values are handled by using the ML method (FIML). Estimates are from linear models with covariates and constants included. *p*-values reported in table are for effects being equal across subgroups (for Background or as indicated for Maternal Ed).

* $p < 0.05$.

** $p < 0.01$.

[a] Low (primary school); low–mid (high school, vocational education); high–mid (e.g., professional BA such as teacher); high (bachelor's degree and advanced university education).

for the hierarchical structure of the dataset. The regression model is specified as:

$$dy_i = \beta_0 + \delta I_i + \boldsymbol{x}'_i\beta + \varepsilon_i$$

where dy_i denotes the change in the outcome of interest for child i and δ represents the intervention effect. Hence, the model describes the determinants of the *change* in outcome from the pretest to the posttest. To increase precision of the estimates, we include child and family covariates represented by the vector $\boldsymbol{x}_i$. The included covariates are gender, age, mother's and father's education, ethnic background, employment status (no employment), and family status (married, cohabitating, or single parent), as shown in Table 1. Results are similar without the inclusion of these covariates (and municipality fixed effects). Intraclass correlations (ICCs) provide evidence for variation in the outcome measures both within and between childcare centers and classrooms. At the childcare center level, ICCs ranged from 0 on language use to 0.09 on math language and at the classroom level, ICCs ranged from 0.16 on language use to 0.26 on math language. These values indicate that the majority of variance was between children, with fractions ranging from 0.74 on math language to 0.84 on language use. The analyses were also run using hierarchical linear modeling, and results were similar to those presented here. Missing data was handled by the maximum likelihood method (FIML), while a sensitivity analysis showed that multiple imputation yielded similar results.

Subgroup analyses were conducted using the same model, but with interaction terms added for the interaction between the intervention indicator I_i and the indicator variables for the relevant subgroups. The model with interaction terms was specified such that we obtained separate estimates for the treatment effect for each subgroup (reported in Table 4).

As many ECE programs have been reported to vary substantially in their provision of high-quality adult–child interactions, additional analyses were conducted to examine the potential interaction between treatment effects and baseline center process quality. Analyses of the potential moderating effect of classroom quality were conducted by adding the EBS and ESL scores into each model as a separate variable and as an interaction term with the intervention indicator (using a two-level hierarchical linear model with children nested in classrooms to capture that we are using a classroom-level variable). These analyses were only conducted for the children in classrooms for which the CLASS was administered.

8. Results

8.1. Main effects of "Play and Learn"

To address the first research, we examined the extent to which the intervention had positive effects on children's school readiness outcomes as compared to the control group. Cohen's *d* effect size estimations based on the results from the regression models appear in Table 4 for each of the language, math and social–emotional skills. As can be seen, the intervention "Play and Learn" had significantly positive effects on children's skills in the four areas examined: general vocabulary ($d = 0.27$), language use ($d = 0.19$), math language ($d = 0.80$), and numeracy ($d = 0.55$). Therefore, the results show the intervention to have significant positive effects on all preacademic outcomes ranging from small in size for the language outcomes and medium to large in size for the two math outcomes. Children receiving the "Play and Learn" intervention did not improve skills relative to the control group on measures of social–emotional skills (self-regulation & cooperation, $d = 0.12$; empathy $d = -0.00$), that is, there were no spill-over effects. However, the intervention did not have a negative impact on social–emotional skills either, even though they were not targeted in the intervention.

8.2. Differential effects of "Play and Learn"

Subgroup analyses were conducted to examine treatment effects for Danish versus non-Western immigrant children (i.e., children starting to acquire Danish as a second language) and for children across various SES strata based on maternal education (see Table 4). For the subgroup analyses, we use information from

Table 5
Descriptive information for the CLASS toddler ($N = 97$ classrooms).

	Control			Intervention		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Emotional and Behavioral Support (EBS)	5.71	.56	4.2–6.7	5.73	.51	4.7–6.8
Engaged Support for Learning (ESL)	3.08	.67	2.0–4.8	2.88	.50	1.9–4.2

Statistics Denmark on ethnic background for the children and their mother's education. We distinguish between Danish children (children with at least one parent born in Denmark and having Danish citizenship) and immigrant children (children with none of their parents being born in Denmark and having Danish citizenship; the child may has been born in Denmark or abroad). We only consider immigrant children from non-Western countries in the subgroup analysis as there were too few children from Western countries. Mother's education is divided into four different categories.

For the Danish and non-Western groups, we find significant treatment effects for four outcomes (vocabulary $d = 0.29$, language use $d = 0.23$, math language $d = 0.88$, and numeracy $d = 0.61$) for Danish children, whereas no significant effects were found for non-Western children ($d = -0.17$–0.15). However, the differences between Danish and non-Western children were not significant for vocabulary and language use ($p = 0.535$ and 0.067, respectively). It should be noted that there was a high degree of concentration of children of non-Western background in classrooms. Almost two-thirds of all classrooms did not have any children of non-Western background, whereas in 8% of the classrooms at least 50% of the children had non-Western background.

Subgroup analyses were also conducted to examine treatment effects across SES strata, based on grouping of maternal education into four groups. For children of low–mid and mid–high-educated mothers, we find strong and significant effects for three outcomes (vocabulary $d = 0.23$–0.37, math vocabulary $d = 0.69$–0.79) and numeracy $d = 0.44$–0.53), whereas only children with high-educated mothers show significant improvement in all four outcomes (vocabulary $d = 0.35$, language use $d = 0.39$, math vocabulary $d = 1.15$, and numeracy $d = 0.77$). For children of low-educated mothers, the treatment only has significant effects for math language ($d = 0.56$) and numeracy ($d = 0.49$). We only find significant differences between children of low- and high-mid educated mothers on language use and empathy ($p = 0.042$–0.046), whereas children of high educated mothers score significantly higher on vocabulary, language use and math language compared to children of low-educated mothers ($p = 0.001$–0.044).

We also investigated whether the intervention effects were conditioned on children's initial skill levels by conducting an analysis with baseline skills (pretest assessments) as a moderator of the effects. This analysis indicated that for three outcome measures, language use, math language and numeracy, there was a significantly positive moderating effect of baseline skills. However, further analysis revealed that the subgroup differences in treatment effects were persistent even when baseline skills were included as a moderator. This additional analysis suggests that the intervention may have been especially beneficial for children from advantaged families. Maternal education and initial skill levels were uncorrelated, whereas children of non-Western background had lower maternal education and initial skill levels (correlations in the range of .11–.16).

8.3. Intervention effects and childcare quality

To examine the potential interaction between treatment effects and baseline center process quality, the CLASS-Toddler data for each program was considered. Table 5 provides descriptive information for the CLASS-Toddler. As can be seen, the mean score of Emotional and Behavioral Support (EBS) is in the mid–high range (5.76), whereas the mean score of Engaged Support for Learning (ESL) is in the low–mid range (2.96) but with substantial variation between classrooms. The interaction between treatment and classroom process quality, which is shown in Table 6, was only significant in one case, namely between language use (−0.43) and ESL.

Table 6
Interaction between treatment and quality at classroom level ($N = 97$ *classrooms/580 children*).

	Vocabulary	Language use	Math language	Numeracy
Simple main effects	0.38**	0.15	0.70**	0.52**
	(0.09)	(0.10)	(0.16)	(0.16)
Interaction effects				
Emotional and	−0.01	−0.22	0.21	0.30
Behavioral Support	(0.15)	(0.18)	(0.27)	(0.27)
Engaged Support	−0.15	−0.43*	−0.04	0.24
For Learning	(0.16)	(0.17)	(0.36)	(0.35)

Notes. Standard errors in parentheses. Standard errors are adjusted for clustering at the classroom level. Estimates are from two-level hierarchical linear models with constants, main effects and interaction terms included. Estimates shown in the second part of the table are of coefficients to the interaction terms with the treatment.
* $p < 0.05$.
** $p < 0.01$.

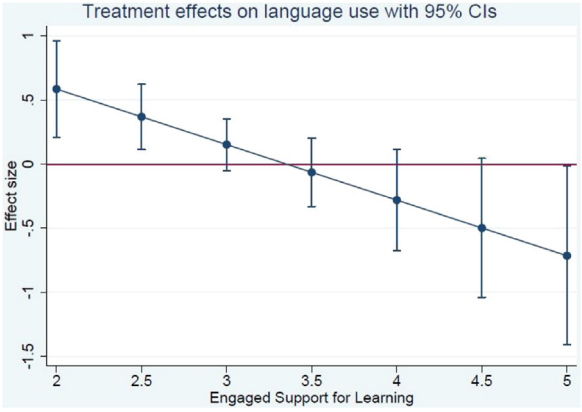

Fig. 2. Treatment effects on language use for different levels of engaged support for learning (ESL).

Based on visualizations of the treatment effect for different levels of ESL for language use, as shown in Fig. 2, we find that the treatment effect on language use is only significantly positive for children in classrooms with initially lower levels of ESL, whereas the effect becomes smaller and insignificant for initial higher levels of ESL. To visualize the estimated coefficient of the interaction term (reported in Table 6), Fig. 2 shows the estimated treatment effect as a function of the ESL scores (i.e., as conditional marginal effects). The results therefore indicate that the intervention effect on children's language use is stronger in classrooms with initially lower levels of ESL in comparison to classrooms with initially higher ESL, indicating that the added value of the intervention is largest for classrooms scoring the lowest.

9. Discussion

Skill begets skill as brains are built from the bottom up; therefore, the earlier children are provided enriching environments that explicitly promote learning of critical skills, the more likely they will develop the foundational skills that underlie future schooling success (Phillips et al., 2017). At the same time, providing high-quality early education programs to young children allows parents to stay in the work force and contribute economically to society (Baker et al., 2008). With these two circumstances well-established, a pressing challenge many local authorities face is how to ensure that teachers who care for very young children are able to stimulate early growth in skills, including language and math. This study, involving 1116 18-to-36-months old children enrolled in the near-universal Danish child-care system, represents one of the first large-scale RCTs focused on increasing the skills of infants and toddlers in ECE settings.

9.1. Main impact of the intervention

The first research question asked to what extent a cost-efficient intervention improved children's language, math language, and math skills. The results of the study showed that the "Play and Learn" intervention had a significant short-term impact on all the targeted language, math language and numeracy skills with overall effect sizes ranging from 0.19 to 0.80, corresponding to medium to large effects. While numerous studies have shown that similar approaches can improve early education outcomes for four-to-five-year old children (Chambers et al., 2016; Duncan & Magnuson, 2013; Shonkoff et al., 2016), the present study suggests that intentional teaching providing children with more support for learning is effective as measured by teacher ratings when children are as young as 18 months of age, at least for children from middle- to upper-class families. This is particularly important in relation to math language and numeracy, as evidence points to young children's early math competencies as being extremely relevant for their future math and reading achievement (Duncan et al., 2007). It should be noted, though, that a likely reason that "Play and Learn" was particularly beneficial for children's development of math language and numeracy skills is that these skills are seldom taught within business-as-usual toddler (BAU) classrooms in Denmark, although we have no empirical data on math activities in the control group. The null finding for social–emotional skills (which were not targeted in the intervention) suggests that the intervention did not distract teachers' emotional and behavioral support of the children. Teachers were able to maintain the high quality of emotional support that was present prior to the intervention, as documented by the pretest CLASS scores. The present study fundamentally contributes to contemporary issues by showing that a relatively easy-to-implement, yet systematically organized, curriculum can significantly improve the language and math skills of toddlers relative to BAU care. Importantly, the curriculum provided did not increase the amount of time children spent in childcare or the number of adults who care for them. Rather, the curriculum provided teachers with tools – training, sequence and scope, strategies, and materials – by which to be more explicit and intentional in their interactions with children, serving as mechanisms affecting children's language and math development. Therefore, the study also adds additional evidence that leaving teachers with some discretion to implement programs can lead to improved outcomes in children but that more is needed to make sure that socially disadvantaged children benefit as well. Suggestively, higher levels of teacher freedom in planning and implementation may have reduced the curriculum complexity and enabled teachers to focus on the most important part of the curriculum, that is, engaging in interactions focused on learning and the sequence and scope may have supported the teachers with age appropriate targets that have stimulated teachers to use richer language. A comparison with the RECC curriculum referred to earlier (Landry et al., 2014) illustrates this last point. The most significant difference of "Play and Learn" compared to the RECC curriculum, is that "Play and Learn" follows a sequence and scope for language and math targets, whereas the RECC is only implementing a sequence and scope for social–emotional skills. Interestingly, in the "Play and Learn" intervention, significant impacts were only found for preacademic skills, whereas in the RECC intervention, significant impacts were only found for social–emotional skills. Note that the intervention dosage is much smaller for "Play and Learn" compared to the RECC, and still intervention effects were stronger for pre-academic skills. This points to the importance of providing a sequence and scope for teachers relative to lesson plans and aligned activities, possibly reflecting the fact that teachers need the more specific information about how children develop skills at specific ages that is provided by the sequence and scope, whereas the aligned activities seem less important. However, it should be noted, that the educational level of Danish teachers is higher than in the US and some other settings as well, with 60% of teachers having a pedagogical bachelor degree. Furthermore, this study shows that the process quality of the Danish toddler classrooms was in the mid–high range for Emotional and Behavioral Support and in the lower end for Engaged Support for Learning reflecting a higher level of process quality compared to the US and some other OECD countries (Burchinal et al., 2015; Slot, 2018). The higher educational level may have served as a necessary background for the teachers to use the sequence and scope and the supportive tools to be more explicit and intentional in their interactions with children. It is therefore unclear if the same intervention effects can be found in ECE settings where teachers are less educated.

9.2. Child and family characteristics and intervention effectiveness

The second research question focused on the extent to which the impact of "Play and Learn" was conditional on child- and parental background characteristics, namely language background and family SES. By focusing attention to subgroups within a single, large experiment, we found that middle- to upper class children seemed to especially benefit from the intervention, which is in contrast to findings from some US studies (Phillips et al., 2017). A further analysis of whether the intervention effects were conditioned on children's initial skill levels revealed that the subgroup differences in treatment effects were persistent even when baseline skills were included as a moderator suggesting that the intervention may have been especially beneficial for children from advantaged homes. Weaker and in many cases non-significant results were found for children of non-Western origin and children with mothers with low education (primary school only). As all dual language learners are tested in Danish (their second language), it may not be surprising that this group of children did not improve on the more advanced language measures (language use and math language), but there were no significant effects for these children on basic vocabulary either. However, dual language learners with non-Western background in Denmark typically only meet the majority language (Danish) when they enter childcare and as dual language learners typically experience an initial silent period in their second language acquisition, thus longer time may be needed for effects to emerge (Krashen, 1985). Similarly, children of low educated mothers benefited less from the intervention on vocabulary and language use. There was no indication of differential exposure to the intervention as a result of socioeconomic factors suggesting that children of low-educated

mothers may need more exposure to experience positive effects (Wen, Leow, Hahs-Vaughn, Korfmacher, & Marcus, 2012). Other potential explanations are that the type of activities that teachers chose to carry out may potentially have been more interesting and engaging for more advantaged children or that implicit biases resulted in too low expectations for less advantaged children's ability to learn, which may have caused teachers to implement the intervention less systematically and with lower quality for these children. More specific training of teachers in supporting children of diverse backgrounds and collegial coaching to support reflection of the implementation of the intervention to meet all children's needs may be a way forward to improve outcomes for all children.

9.3. Childcare quality and intervention effectiveness

The last research questions examined the degree to which intervention impacts were influenced by the quality of the childcare prior to the intervention. We only found limited evidence that the existing quality of the classroom as measured by the CLASS Toddler instrument served to moderate intervention effects. For one language domain, language use, the intervention interaction with quality suggests that a curriculum like "Play and Learn" may be even more helpful for teachers who provide lower levels of ESL. This suggests that "Play and Learn" can improve child outcomes in classroom independent of the classroom quality.

9.4. Limitations

The findings of this study are important to advancing investigations of strategies for enhancing the quality and outcomes of toddler care, and is one of the largest causally interpretable studies of its type. While the study findings show the potential value of incorporating intentional teaching via responsive practices and a scope and sequence in such settings, several limitations should be noted. First, the understanding of classroom practices in control settings was only studied via use of a general observational tool and thus relatively little is known about the counterfactual in this study. The study results would have been more strongly contextualized if the counterfactual conditions were better understood. Second, as studies demonstrate that the CLASS has small linkages with child outcomes, it would be important in future studies to consider additional indicators of program quality – over and above CLASS – to investigate interactions between treatment effects and existing center process quality. Third, the assessment of child outcomes relied on teacher report instruments, including the math checklist developed as part of this study. Although teacher-report instruments are often used in studies of ECE practices, we acknowledge that this limits the strength of the conclusions drawn from the study and that standardized assessments administered by a third party would be a valuable addition to estimating treatment effects. Fourth, long-term follow-up data are not available, thus it is unknown if the intervention results will have longer-term benefits for children's development, which is the goal of such work. It is especially important to determine whether the intervention benefits those children for whom early educational experiences may be particularly vital, such as children from low-SES families. Finally, it is currently unclear whether the intervention effects would successfully generalize to other contexts like the US. Structural characteristics like a higher educational level of the Danish teachers and higher teacher–child ratios may have affected the implementation of the open curriculum and more research in other cultural settings is needed.

9.5. Conclusion

With the increased demand for ECE programs due in part to the rise of mothers joining the work force, many countries are faced with a pressing challenge to provide high quality early education that stimulates early growth in children's language and math skills. The present study, conducted within childcare centers in a welfare state, demonstrates the effectiveness of a 20-week curriculum for increasing children's short-term language and math outcomes, even when children are as young as 18 months of age over a relatively short period of time; importantly, the curriculum allows teachers with some discretion to implement the intervention and relies on the use of responsive interactions with children. In particular, the large effects on math are especially important as they demonstrate that it is feasible to increase math-related vocabulary and early numeracy skills in toddlers.

Authors' contribution

Dorthe Bleses: Conceptualization, methodology, writing – original draft preparation, project administration, funding acquisition; *Peter Jensen*: Methodology, data curation, writing – original draft preparation. *Pauline Slot*: Conceptualization, writing – reviewing and editing. Laura Justice: Conceptualization, writing – reviewing and editing.

Acknowledgements

This research was supported from a grant by TrygFonden (project number 21865). We wish to thanks the participating municipalities, teachers, and children.

Appendix A

Tables A1 and A2.

Table A1
Comparison of attriters and non-attriters for the total aample and separately for control and intervention groups.

Variable	Total sample *p*-values	Control *p*-values	Intervention *p*-values
Parent characteristics			
Mother low education (0/1)	0.49	0.85	.37
Mother low/mid education (0/1)	0.68	0.69	.82
Mother high-mid education (0/1)	0.22	0.54	.24
Mother high education (0/1)	0.31	0.35	.60
Mother missing data on education (0/1)	0.13	0.84	.08
Father low education (0/1)	0.14	0.04*	.90
Father low/mid education (0/1)	0.59	0.73	.63
Father high-mid education (0/1)	0.78	0.14	.28
Father high education (0/1)	0.26	0.81	.19
Father missing data on education (0/1)	0.54	0.79	.57
No employment (income base) (0/1)	0.47	0.55	.82
Parent cohabiting (0/1)	0.20	0.78	.13
Parents married (0/1)	0.35	0.62	.40
Living with single parent (0/1)	0.69	0.64	.24
Child characteristics			
Boy (0/1)	0.08	0.03*	.67
Non-Western origin (0/1)	0.29	0.82	.20

Notes: Education levels: low (primary school); low–mid (high school, vocational education); high–mid (e.g., professional BA such as teacher); high (BA and advanced university education). The tests are for equality of means in the group of attriters and the group of non-attriters. The *p*-values are from regressions of row variables on an attrition indicator variable with standard errors adjusted for clustering at the childcare center level.

* $p < 0.05$.

Table A2
Characteristics of children and parents in the initial sample ($N = 1508$ children).

Variable	Control	Intervention	Balancing tests
	Mean	Mean	*p*-values
Parent characteristics			
Mother low education (0/1)	0.11	0.12	.51
Mother low/mid education (0/1)	0.27	0.27	.93
Mother high-mid education (0/1)	0.29	0.28	.61
Mother high education (0/1)	0.25	0.24	.75
Mother missing data on education (0/1)	0.08	0.09	.35
Father low education (0/1)	0.16	0.13	.46
Father low/mid education (0/1)	0.35	0.36	.64
Father high-mid education (0/1)	0.20	0.19	.56
Father high education (0/1)	0.21	0.23	.67
Father missing data on education (0/1)	0.09	0.09	.95
No employment (income base) (0/1)	0.13	0.10	.37
Parent cohabiting (0/1)	0.28	0.29	.78
Parents married (0/1)	0.56	0.56	.92
Living with single parent (0/1)	0.14	0.14	.82
Child characteristics			
Boy (0/1)	0.53	0.52	.96
Non-Western origin (0/1)	0.08	0.13	.16
Age (months)	25.8	25.4	.27

Notes: Education levels: low (primary school); low–mid (high school, vocational education); high–mid (e.g., professional BA such as teacher); high (BA and advanced university education). Balancing tests are for equality of means in the control and intervention group. The *p*-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the childcare center level. $^{*}p < 0.05$.

References

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*, 7–39.

Baker, M., Gruber, J., & Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of political Economy*, *116*, 709–745.

Bauchmüller, R., Gørtz, M., & Rasmussen, A. W. (2014). Long-run benefits from universal high-quality preschooling. *Early Childhood Research Quarterly*, *29*, 457–470.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, *107*, 251–271.

Bleses, D., Højen, A., Dale, P. S., Justice, L. M., Dybdal, L., Piasta, S., . . . & Haghish, E. (2018). Effective language and literacy instruction: Evaluating the importance of scripting and group size components. *Early Childhood Research Quarterly*, *42*, 256–269.

Bleses, D., Højen, A., Justice, L., Dale, P., Dybdal, L., Piasta, S., Markussen-Brown, J., Clausen, M. C., & Haghish, E. F. (2018). The effectiveness of a large-scale early childhood language and pre-literacy intervention: The SPELL randomized-controlled-trial in Denmark. *Child Development*, *89*, e342–e363.

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., Nielsen, H., Sehested, K., & Sjoe, N. (2016). *Børns tidlige udvikling og læring, Early childhood development in Danish childcares*. For Ministry of Children, Education and Gender equality.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*, 1461–1476.

Broekhuizen, M., Leseman, P., Moser, T., & van Trijp, K. (2015). Stakeholders Study. Values, beliefs and concerns of parents, staff and policy representatives regarding ECEC services in nine European countries: First report on parents. *EU CARE Project*.

Burchinal, M., Magnuson, K., Powell, D., & Soliday-Hong, S. (2015). Early childcare and education. In M. Bornstein, R. Lerner, & T. Leventhal (Eds.), *Handbook of child psychology and developmental science* (Vol. IV) (pp. 223–267). New York, NY: Wiley Press.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*, 166–176.

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.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, *333*(6045), 968–970.

Crawford, A., Zucker, T., Van Horne, B., & Landry, S. (2017). Integrating professional development content and formative assessment with the coaching process: The Texas School Ready Model. *Theory Into Practice*, *56*, 56–65.

Dale, P. S., Price, T. S., Bishop, D. V., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, and Language, and Hearing Research*, *46*, 544–560.

Denham, S. (2006). Social–emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, *17*, 57–89.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . & Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, *27*, 109–132.

Fenson, L., Marchman, V., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories. Users guide and technical manual* (2nd ed.). Baltimore: Paul H. Brookes Publishing Co.

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. In *What works clearinghouse*.

Gupta, N. D., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, *94*, 30–43.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . & Scott-Little, C. (2012). A course on effective teacher–child interactions effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, *49*, 88–123.

Hatfield, B. E., Burchinal, M., Pianta, R., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, *36*, 561–571.

Helmerhorst, K. O., Riksen-Walraven, J. M. A., Fukkink, 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.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*, 4–14.

Justice, L. M., Logan, J. A., Purtell, K., Bleses, D., & Højen, A. (2017). Does mixing age groups in early childhood education settings support children's language development? *Applied Developmental Science*, *23*, 214–226.

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and literacy curriculum supplement for preschoolers who are academically at risk: A feasibility study. *Language, Speech, and Hearing Services in Schools*, *41*, 161–178.

Jørgensen, R., Dale, P., Bleses, D., & Fenson, L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of child language*, *37*, 419–428.

King, E. K., Pierro, R. C., Li, J., Porterfield, M. L., & Rucker, L. (2016). Classroom quality in infant and toddler classrooms: Impact of age and programme type. *Early Child Development and Care*, *186*, 1821–1835.

Krashen, S. (1985). *The input hypothesis*. Beverly Hills: Laredo Publishing Company.

La Paro, K. M., Hamre, B. K., & Pianta, R. C. (2012). *Classroom assessment scoring system (CLASS) manual, toddler*. Paul H. Brookes Publishing Company.

La Paro, K. M., Williamson, A. C., & Hatfield, B. (2014). Assessing quality in toddler classrooms using the CLASS-Toddler and the ITERS-R. *Early Education and Development*, *25*, 875–893.

Landry, S. H., Zucker, T. A., Taylor, H. B., Swank, P. R., Williams, J. M., Assel, M., . . . & 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.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*, 69–92.

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, *24*, 305–337.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*, 97–115.

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*.

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*, *260*, 66–88.

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social–emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education & Development*, *24*, 1000–1019.

Norris, D., & Horm, D. (2015). Teacher interactions with infants and toddlers. *Young Children*, *70*, 84–91.

Nøhr, K., Dalsgaard, C., Kloppenborg, H., Meldgaard, J., & Bækgaard, M. (2012). *Sådan passer kommunerne børn, This is how the municipalities serve children*. KORA (Det Nationale Institut for Kommuners og Regioners Analyse og Forskning).

OECD. (2017). *OECD family data base*.. Retrieved from http://www.oecd.org/social/family/database.htm

Pianta, R., La Paro, K. M., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS) manual, Pre-K*. Baltimore, MD: Paul H. Brookes Publishing.

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. A consensus statement*. pp. 19–30. Issues in pre-kindergarten programs and policy.

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.

Shonkoff, J., Levitt, P., Fox, N., Bunge, S., Cameron, J., & Duncan, G. (2016). *From best practices to breakthrough impacts: A science-based approach for building a more promising future for young children and families*. Cambridge, MA: Harvard University, Center on the Developing Child.

Sjoe, N. M., Bleses, D., Dybdal, L., Nielsen, H., Sehested, K. K., Kirkeby, H., . . . & Jensen, P. (2017). Measurement properties of the seam questionnaire using Rasch analysis on data from a representative Danish sample of 0-to 6-year-olds. *Journal of Psychoeducational Assessment*, *37*, 320–337.

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. http://dx.doi.org/10.1787/edaf3793-en

Slot, P. L., Bleses, D., Justice, L. M., Markussen-Brown, J., & Højen, A. (2018). Structural and process quality of Danish preschools: Direct and indirect associations with children's growth in language and preliteracy skills. *Early Education and Development*, *29*, 581–602.

Statistics Denmark. (2014). *Indvandrere i Danmark 2014, Immigrants in Denmark*. Copenhagen, Denmark: Statistics Denmark.

Squires, J. (2014). *Social–emotional assessment/evaluation measure (seam)*. Paul H. Brookes Publishing.

Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher–child interactions in toddler child care. *Early Education and Development*, *20*, 285–304.

Vach, W., Bleses, D., & Jørgensen, R. N. (2010). Construction of a Danish CDI short form for language screening at the age of 36 months: Methodological considerations and results. *Clinical Linguistics & Phonetics*, *24*, 602–621.

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.

Vandell, D., Burchinal, M., & Pierce, K. M. (2016). Early child care and adolescent functioning at the end of high school: Results from the NICHD Study of Early Child Care and Youth Development. *Developmental Psychology*, *52*, 1634–1645.

Wen, X., Leow, C., Hahs-Vaughn, D. L., Korfmacher, J., & Marcus, S. M. (2012). Are two years better than one year? A propensity score analysis of the impact of Head Start program duration on children's school performance in kindergarten. *Early Childhood Research Quarterly*, *27*, 684–694.

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.