CHILD DEVELOPMENT

Child Development, xxxx 2017, Volume 00, Number 0, Pages 1–22

The Effectiveness of a Large-Scale Language and Preliteracy Intervention: The SPELL Randomized Controlled Trial in Denmark

Dorthe Bleses and Anders Højen
Aarhus University

Laura M. Justice
The Ohio State University

Philip S. Dale
University of New Mexico

Line Dybdal
Rambøll Management Consulting

Shayne B. Piasta
The Ohio State University

Justin Markussen-Brown and Marit Clausen
University of Southern Denmark

E. F. Haghish
University of Southern Denmark and University of Freiburg

The present article reports results of a real-world effectiveness trial conducted in Denmark with six thousand four hundred eighty-three 3- to 6-year-olds designed to improve children's language and preliteracy skills. Children in 144 child cares were assigned to a control condition or one of three planned variations of a 20-week storybook-based intervention: a base intervention and two enhanced versions featuring extended professional development for educators or a home-based program for parents. Pre- to posttest comparisons revealed a significant impact of all three interventions for preliteracy skills (= .21–.27) but not language skills (d = .04–.16), with little differentiation among the three variations. Fidelity, indexed by number of lessons delivered, was a significant predictor of most outcomes. Implications for real-world research and practice are considered.

Young children's language and preliteracy skills are important foundations to their future reading skill in adolescence and adulthood (e.g., Verhoeven & Van Leeuwe, 2008). Not surprisingly, children who arrive to formal reading instruction with well-developed language and preliteracy skills have elevated academic trajectories compared to children who exhibit lags in these areas (National Early Literacy Panel [NELP], 2008). Thus, there is increasing interest in identifying children who exhibit lags in language and preliteracy development during the preschool years and providing them with early educational opportunities to improve these skills. To this end, a number of studies have tested interventions designed to boost children's early language and preliteracy skills (NELP, 2008). Many of these interventions have resulted in benefits to young children that are considered to be of practical as well as statistical significance based on consideration of effect-size estimates (e.g., Lipsey et al., 2012). Furthermore, some work finds that improvements in early language and preliteracy skills have effects that last into the early and later elementary grades (Piasta, Justice, McGinty, & Kaderavek, 2012; Whitehurst et al., 1999).

The present study represents an important complement to the growing corpus of efficacy studies investigating the effects of language and preliteracy interventions for young children, as it examines intervention effects when implemented at scale

This research was supported from a grant by The Danish Council for Strategic Research (No. 11-115756). Additional support for books used in the intervention came from two publishers, Gyldendal and Forlaget Torgard.

Minor portions of the Introduction and Method sections of this article are adapted from the study protocol published by Bleses, Højen, Dybdal, Dale, Justice, Piasta, Markussen-Brown, Clausen, Haghish, Aktürk Ari, Andersen and Vach (2014).

Correspondence concerning this article should be addressed to Dorthe Bleses, Trygfonden's Centre for Child Research and School of Communication and Culture, Aarhus University, Fuglesangs Allé 4, building 2630, 8210 Aarhus V, Denmark. Electronic mail may be sent to bleses@econ.au.dk.

 0009-3920/2017/xxxx-xxxx
DOI: 10.1111/cdev.12859

in an unselected heterogeneous sample within real-world conditions. The study was implemented across Denmark, a northern European context in which child care programs typically provide little systematic focus on developing children's language and preliteracy skills; rather, Danish child care programs emphasize socialization and play. As there is a growing interest in understanding benefits of early language and preliteracy interventions on immigrant children and dual language learners (DLLs; see Farver, Lonigan, & Eppe, 2009) in the northern European context, this effectiveness study included exploration of whether intervention impacts were associated with key background variables of children, including socioeconomic status (SES) backgrounds and dual language learning (i.e., children of nonnative Danish background). Results of this study are informative to considering whether efficacious interventions can be successfully taken to scale to provide benefits for all young children.

Implementing Systematic Language and Preliteracy Interventions for Young Children

A critical element of efficacious early childhood language and preliteracy interventions is the provision of systematically organized and explicit instruction that targets children's development of skills that are predictive of future reading achievement (NELP, 2008). Systematically organized instruction utilizes learning sequences that are guided by a rigorous understanding of how children learn, whereas explicit instruction clarifies to children what they have to attend to within an activity by orienting them toward the goals of the activity (Mesmer & Griffith, 2005). Targeted skills in language and preliteracy interventions often include phonological awareness, print awareness, and oral language skills, as such skills have well-documented relations with future reading achievement (Catts, Herrera, Nielsen, & Bridges, 2015). Manualized language and preliteracy interventions that yield positive effects on children's skills typically feature these elements in some combination (Justice, Kaderavek, Fan, Sofka, & Hunt, 2009; Piasta et al., 2012), resulting in improvements in children's language and preliteracy skills by about 0.25–0.87 *SD* units (see NELP, 2008).

To date, the majority of language and preliteracy interventions have been assessed with samples of children exhibiting risk; these include children raised in low-SES homes (e.g., Lonigan & Phillips, 2016; Piasta et al., 2012) and, increasingly, children who are DLLs. This focus is justified by substantial evidence documenting that these children consistently exhibit less-developed language and preliteracy skills during the preschool years than their peers (Hoff, 2013). However, relatively few studies have examined whether there are differential intervention effects for children as a function of SES and DLL status, thus the extant research does not clarify whether children with such risk factors respond differently to language and preliteracy interventions than their peers. Indeed, the available literature is somewhat mixed. In a meta-analysis of vocabulary interventions, Marulis and Neuman (2013) found that high-SES children benefited significantly more from vocabulary interventions than low-SES children. In contrast, the meta-analysis of the NELP (2008) reported finding larger effects for preliteracy outcomes for low-SES children compared to high-SES children. Given such findings, it is not clear whether children exhibiting SES-related risk factors require different types of language and preliteracy intervention, such as more intensive offerings, than children without such risk factors.

The same can be stated with respect to DLL children, especially given that relatively few studies have systematically examined language and preliteracy interventions for native versus nonnative (DLL) children. The meta-analysis of the NELP (2008) reported higher effect sizes for Latino than Caucasian or African American children, suggesting that DLLs may benefit more from language and preliteracy interventions compared to not DLLs (assuming the Latino children included were, in fact, DLLs). However, the differences were not significant, possibly because too few studies could be analyzed. More recently, a systematic review evaluated the impact of early education programs and practices for 3- to 5-year-old Spanish English DLLs' development (Buysse, Peisner-Feinberg, Páez, Hammer, & Knowles, 2014); the review included four studies focused specifically on language and preliteracy interventions (all using English as the language of instruction), one of which featured a randomized controlled trial (RCT) design. Across the four studies, the results suggested that DLL children's language and preliteracy development were improved as a result of intervention exposure, similar to findings seen in other works involving DLLs acquiring languages like Arabic, Somali, and Turkish as their first language and either Dutch or English as their second language (Dockrell, Stuart, & King, 2010; van Tuijl, Leseman, & Rispens, 2001; Verhallen & Bus, 2010). Interestingly, none of these studies have explored intervention effects in a

single study for both non-DLL and DLL children; therefore, we cannot determine whether DLLs' respond similarly to exposure to language and preliteracy intervention, beyond drawing comparisons across studies. Research that assesses whether DLLs and non-DLLs derive similar benefits from language and preliteracy interventions would be highly informative to both theory and practice.

In the present study, we examine effects of a systematic, explicit language- and preliteracy-focused intervention for a general population of preschool-aged children, as well as effects for two subsamples of children experiencing risk: children from lower SES homes and children who are DLLs. By including attention to these subgroups within a single, large experiment, we can assess whether intervention effects are conditional upon presence, or lack thereof, of these common risk factors among children. Although there is a clear societal interest in understanding intervention effects for children at risk, there are also important reasons to explore the effects of language and preliteracy interventions for a broader population of children. First, our ability to identify children who will exhibit poor reading in their future is too imperfect to identify all children in need for early intervention (e.g., Bornstein, Hahn, Putnick, & Suwalsky, 2014). Second, the growing demand for creating new technologies and innovation has led to a global competition for highly skilled, literate employees. Finally, within Denmark, there are concerns about the relatively modest levels of reading achievement observed among students despite very high levels of public investment in schooling (Organisation for Economic Development & Cooperation [OECD], 2010).

Enhancing Intervention Effects Through Teacher and Parent Supplements

The typical approach to investigating the effects of early language and preliteracy interventions involves child care providers or early educators implementing experimental programs and practices within the classroom milieu (e.g., Bierman et al., 2008; Fischel et al., 2007; Justice et al., 2009; Landry, Swank, Anthony, & Assel, 2011). Although intervention effects for phonological awareness interventions are relatively high (~.8), effects on other preliteracy (e.g., print awareness) and language skills are more modest (~.3 to .4; see NELP, 2008). As an avenue for increasing these effects, researchers have increasingly turned to exploring whether increased teacher training, or professional development (PD), can improve results for children, given evidence that PD can improve teachers practices and, to some extent, child outcomes (e.g., Landry, Anthony, Swank, & Monseque-Bailey, 2009; Powell, Diamond, Burchinal, & Koehler, 2010). Researchers have also explored whether involving parents as intervention agents can also enhance effects (Fey, Cleave, Long, & Hughes, 1993; Justice, Logan, Kaderavek, & Dynia, 2015).

In relation to PD, many interventions are based on the rationale that improving educators' knowledge about language and preliteracy facilitation strategies, often combined with coaching of educators' practice during intervention, will lead to improved fidelity of the intervention and, in turn, improved child outcomes. In part, the focus on improving educators' fidelity stems from studies showing relatively low levels of fidelity when educators implement early language and preliteracy interventions (Pence, Justice, & Wiggins, 2008). A majority of studies have investigated effects of different forms of PD (see Sheridan, Edwards, Marvin, & Knoche, 2009), whereas fewer have compared the effect of a curriculum by itself as compared to the value-added benefit of providing educators with additional PD. An exception is work by Lonigan, Farver, Phillips, and Clancy-Menchetti (2011), who evaluated the effect of the Literacy Express Preschool Curriculum as coupled with two PD components: (a) workshop training and (b) mentoring (Lonigan et al., 2011). Interestingly, there was only one significant difference in child outcomes to favor the more intensive PD model involving mentoring (see also Assel, Landry, Swank, & Gunnewig, 2007). However, other evidence has shown that ongoing provision of intensive PD can enhance the quality of systematic and explicit interventions (Hamre et al., 2012; Wasik & Hindman, 2011), such PD has focused training teachers to detect effective adult–child interactions via video analyses (Hamre et al., 2012), increasing teacher knowledge of children's language and preliteracy development Landry et al., 2009, 2011), and offering teachers opportunities to engage in ongoing collegial discussion (Buysse, Sparkman, & Wesley, 2003). Although such results are promising, these PD approaches have only been evaluated for their impacts on at-risk children, primarily those from low-SES homes, and never as an add-on component to a curriculum. In the present study, we determined whether coupling intensive, supportive PD of child care providers with implementation of a systematic, explicit intervention would improve children's outcomes.

In addition to providing support to teachers via provision of PD, a large body of work has argued the importance of supporting parents of young children to implement language and preliteracy activities in the home environment (Bierman, Welsh, Heinrichs, Nix, & Mathis, 2015; Fey et al., 1993; Justice et al., 2015; van Steensel, McElvany, Kurvers, & Herppich, 2011). Theoretically, supporting parents to engage in learning-related activities at home can help to empower parents and strengthen home–school relations (Rimm-Kaufman & Pianta, 2000). However, parental implementation of language and preliteracy activities in the home environment can also provide an additional dosage of learning-related opportunities. Bierman et al. (2015) recently found that a home-visiting program that helped low-SES parents to implement language and preliteracy activities in the home elevated the effects of teachers' implementation of a comprehensive preschool curriculum.

However, the research to date shows some mixed findings regarding the value of aligned implementation by parents and teachers of early language and preliteracy programming. Manz, Hughes, Barnabas, Bracaliello, and Ginsburg-Block (2010) conducted a meta-analysis involving 14 studies relevant to this topic, reporting that effects of aligned interventions on children's language and preliteracy skills was significantly lower than that of parent-only interventions (d = .13 and .47, respectively). Such work suggests that implementation by one provider, in this case the parent, was more powerful than implementation by two (parent and early educator). Justice et al. (2015) reported a similar finding for children receiving a year-long preliteracy intervention in their classrooms; children whose teachers implemented the intervention solo had better outcomes than children whose teachers and parents simultaneously implemented the intervention. These authors suggested that there may be dosage thresholds, whereby exposure to interventions at home and school results in diminishing results over time. However, these studies largely involve at-risk parents (e.g., low educational level) and at-risk children (e.g., developmental disability). In the present study, we examine the potential value-added effect of parent + school coimplementation of a language and preliteracy intervention in a large, unselected sample of children and their families.

Scaling Up Language and Preliteracy Interventions

Efficacy trials, also known as explanatory trials, are distinct from effectiveness trials, also known as pragmatic trials (Schwartz & Lellouch, 1967). Efficacy trials test the effect of an intervention in an ideal, tightly controlled situation, whereas effectiveness studies seek to examine intervention benefits as they occur in real-world conditions; the latter type of study seeks to produce evidence of practical importance that is relevant to all individuals (e.g., Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). Generally, the efficacy–effectiveness distinction is best viewed as a continuous rather than dichotomous continuum: Efficacy trials focus on homogeneity to obtain high internal validity and controlling for as many factors as possible to avoid bias, whereas effectiveness trials focus on maximizing heterogeneity in all aspects to obtain high external validity.

Typically, efficacy studies take place in a resource-intensive "ideal" setting, use relatively small and homogenous participant samples, have strong requirements for provider expertise, evaluate highly standardized and strictly implemented interventions, employ intensive resources in order to maximize compliance, and utilize double- and triple-blinded designs whenever possible. Effectiveness trials, on the other hand, typically take place in a real-world everyday setting, use large and heterogeneous samples, rely on representative usual providers, implement interventions with no or limited reinforcement or compliance support beyond normal practice, and accept that blinded designs are often not possible (Singal, Higgins, & Waljee, 2014). A feature of effectiveness studies is that they often have high levels of missing data as well as participant attrition, given their implementation in real-world settings (O'Brien et al., 2012). Thus, to ensure that effects are unbiased even when attrition is high, effectiveness trials often seek to include an intent-to-treat framework, in which outcomes for all participants, even those who attrite and/or who experience low levels of the treatment, are included in the assessment of treatment outcomes. An additional characteristic of effectiveness studies is that implementers' fidelity to the intervention tends to be lessened relative to the more highly controlled efficacy trials. Therefore, measuring fidelity of an intervention is an essential component of effectiveness trials, and analyses of the relations between implementation fidelity and treatment outcomes can be helpful in exploring key mechanisms of the intervention when implemented widely. As intervention effects are often diluted as a result of lower fidelity (Hulleman & Cordray, 2009), effectiveness trials must be large enough to have sufficient power to

detect relatively small effects and to address issues of attrition.

To date, the vast majority of rigorously conducted studies of language and preliteracy interventions have been efficacy studies with extensive involvement of researchers in the intervention; therefore, our understanding of treatment effects for children is constrained by those circumstances (NELP, 2008). The effectiveness of such interventions as implemented under real-life circumstances and at a large scale is not clear, and this is particularly true outside the context of the United States, in which the majority of such studies have been conducted. A comprehensive literature review of language and literacy interventions published in peer-reviewed journals was done by our team in preparation for this trial. To be included, studies must have evaluated a specific language and literacy intervention that target the children directly, or focused on PD of teachers or parents and included an endpoint measure of children's language, preliteracy, and/or literacy skills. The design of the studies had to be either RCTs or quasi-experimental designs that matched groups on language skills and background factors (Bleses, Dale, Højen, Justice, Jensen, & Andersen, 2017). Our review of the literature identified 141 such studies, but only six large-scale ($n > 500$ children) RCTs of child-care-based language and preliteracy interventions that may be conceptualized as effectiveness trials. These interventions were each carried out in real-life settings by representative usual providers with no or limited reinforcement of fidelity and with large samples of children to increase power. Details of these studies appear in the Appendix. Two studies evaluated the effects of language and/or preliteracy curricula (Fischel et al., 2007; Neuman, Newman, & Dwyer, 2011), one study evaluated a curriculum in combination with PD (Lonigan et al., 2011), and three studies evaluated various PD approaches (Landry et al., 2009; Powell et al., 2010; Wasik & Hindman, 2011). The sample sizes varied from 507 to 1,789. Overall, effect sizes of these six effectiveness trials were noticeably lower than those seen in comparable efficacy trials (NELP, 2008) and ranged from negligible (−.03) to medium (.47) in size. Averaged across studies and samples (and without any weighting for sample sizes), the mean effect size for significant contrasts was about .17 for phonological awareness, .30 for alphabet knowledge, and .16 for vocabulary. The results of such studies suggest that early language and preliteracy interventions have positive effects for children when implemented at scale, although such work is conditioned on the context (United States) and the samples (all at risk).

The Present Study and Intervention Overview

The present study was designed to assess the effects of a large-scale RCT conducted across Denmark under real-life conditions; the study involved 6,483 unselected children. As an effectiveness trial conducted outside of the North American context, the present study expands the literature in several ways. First, learning standards within the United States are commonplace, expanding out of the accountability initiatives of the early 2000s. At the federal level, standards are stipulated for the national Head Start program, and as of a decade ago, a majority of states had adopted early learning standards (Scott-Little, Kagan, & Frelow, 2006). Denmark, in contrast, has no learning standards applied to young children, and there is seldom a systematic instructional focus on either language or preliteracy skills during the preschool years. Not surprisingly, the term "child care" is typically used to describe early education settings rather than the term "preschool," as the latter emphasizes the notion of "schooling."

Second, this is the first large-scale effectiveness trial of which we are aware that it involves an unselected heterogeneous sample of preschoolers. This is an important feature, as it allows us to ask whether results often seen for samples of at-risk children are also seen for a general population of children. It may be that targeted early language and preliteracy interventions are only advantageous for children exhibiting risk, yet it is also possible that these are beneficial to all young children. The present study thus includes all children in the enrolled child cares and not only those who are in programs serving high-risk children. Of note, the overwhelming majority of 3- to 5-year-olds in Denmark (approximately 97%) are enrolled in child care (Danish Ministry for Social Affairs, 2015), with among the highest participation rates among the OECD countries.

Third, the universal child care tradition and a large sample size within this study make it possible to investigate potential differential benefits of the intervention for different groups of children, including low-SES children and DLLs. For the latter group, DLLs in Denmark are distinct from DLL children in the United States, as the largest groups of immigrant children originate from Turkey, Lebanon, Pakistan, Iraq, and Somalia (Statistics Denmark, 2014). The term DLL in the present

study refers to children of immigrant parents in Denmark, for whom Danish is a second or later language. Although these risk groups may differ in important ways from those in the United States, concerns about the language and preliteracy achievement of low-SES children and DLL children are shared. Children from low-SES homes in Denmark are performing 0.33 *SD* below their more advantaged peers, and one half of DLL children from a nonwestern background perform more than 1 *SD* below the mean of monolingual children (Bleses, Højen, Jørgensen, Jensen, & Vach, 2010). Interestingly, even for children without risk factors, concerns also exist: Denmark has a lower percentage of readers at the top proficiency level compared to the OECD average (5% vs. 8%), and the proportion of top performers in reading decreased from 2000 to 2009 (OECD, 2010). In this regard, there is significant interest within Denmark that early education and the universal child care system be used as an avenue for improving all children's language and preliteracy skills as a means to enhance eventual educational attainment in school.

To address such needs, this effectiveness trial involved evaluation of three planned variations of an existing intervention, Read It Again-PreK! (RIA; Justice & McGinty, 2013), previously studied in several efficacy trials in the United States (e.g., Justice et al., 2010). According to an analysis of the intervention by the National Center on Quality Teaching and Learning (2015), RIA shows evidence of addressing 14 of the 16 desired components of an effective language and literacy curriculum. RIA features a fourfold scope of instruction targeting development of vocabulary, narrative, print knowledge, and phonological awareness via 23 sequenced objectives. These objectives are organized into sixty 30-min lessons (each lesson targets two objectives), designed to be delivered to a large group teacher-led session twice weekly over an academic year.

RIA was selected for adaptation into a Danish version, SPELL (Structured Preschool Efforts in Language and Literacy), for several reasons. First, it is a wholly manualized intervention that requires no specialized materials for implementation aside from a set of commercially available storybooks. This was important for our large-scale trial, such that we would not have to provision classrooms with the large volume of manipulatives that often are featured in language and preliteracy interventions. Second, RIA features soft-scripted, structured lesson plans for educators to follow. Thus, many educators typically have little difficulty implementing the lessons with high levels of fidelity (Piasta, Justice, McGinty, Mashburn, & Slocum, 2015). Given the large number of educators included in this trial, and the corresponding low levels of training to be provided to educators in using the intervention, this was a desirable characteristic. Finally, RIA is freely available on the Internet, thus having the potential to achieve a large number of end-users at a very low cost (https://earlychildhood.ehe.osu.edu/research/practice/read-it-again-prek/).

RIA was adapted to the Danish context and language by a university-based team in Denmark involving researchers from several different disciplines (linguistics, speech–language pathology, education) as well as the first author of the original RIA. Several key adaptations were made. First, the overall structure of the intervention, involving twice weekly whole-class sessions for 30 weeks, was redesigned to better fit the Danish child care context. The length of the intervention was decreased from 30 to 20 weeks so that the intervention could fit within a 5-month period of implementation. Given that Danish child care providers do not commonly use any structured program or curricula, there was consensus that shortening the intervention was necessary. Furthermore, RIA was modified from a whole-class to small-group intervention, as whole-class instruction is seldom if ever used with young children in Danish child cares. Second, the 23 RIA objectives were translated into Danish with relatively modest changes, given that the objectives in the original English version are not specific to English; rather, these skills represent important universals in language and preliteracy development, such as learning about print directionality and the concept of word (Bialystok, Luk, & Kwan, 2005). Third, 10 Danish-language storybooks were selected to replace the English storybooks in the original version. Books were selected that were amenable for targeting the 23 RIA objectives, such as having interesting words to discuss for vocabulary development and compelling print features to discuss. Input was solicited from a range of individuals on the suitability of these books for repeated readings in child care settings. Fourth, RIA lessons were drafted by the team to feature the Danish storybooks to deliver the 23 objectives. The lessons used the original template of RIA lessons, which feature a structured sequence of *before reading*, *during reading*, and *after reading* discussions to address the lesson's objectives. The lessons were reviewed and piloted by Danish child care providers. Fifth, the lessons as well as other RIA components (described

in Method) were adapted for delivery via an iPad-based digital learning technology, which offered several benefits. First, the use of a digital platform was deemed more cost effective and environmentally friendly than preparing more than 600 versions of a 300-page manual (one per educator). Second, the digital platform allowed us to collect key data informatics regarding implementation, such as the number and length of individual lessons. A comprehensive overview of SPELL is available online (earlychildhood.ehe.osu.edu/files/2016/04/Read_It_Again_PreK_Danish_Version_SPELL.pdf).

As the first-ever evaluation of a structured language and preliteracy intervention within the Danish child care context, our initial aim was to assess effects of SPELL on children's outcomes within the context of minimal implementation supports. That is, we sought to determine whether provision of the intervention to teachers, with modest supports for implementation, would result in effects similar in magnitude to those seen in efficacy studies of this and similar interventions. However, we also sought to understand whether provision of additional supports in the form of more intensive PD for educators or parental involvement of a companion home-based intervention could improve upon the basic, anticipated effects, as might be suggested based on the literature reviewed previously. Thus, we designed and tested two alternatives to the SPELL intervention: (a) SPELL plus extended PD for educators (henceforth SPELL + PD) and (b) SPELL plus a parent-implemented home-based version (henceforth SPELL + HOME). The goal of the former was to improve educators' understanding of the intervention and fidelity of implementation, deemed instrumental in some studies to enhancing children's outcomes (e.g., Landry et al., 2011). The goal of the latter was to build school + home connections and increase children's learning opportunities, substantiated in recent studies of curricular effects in early education settings (Bierman et al., 2015).

In sum, four questions were addressed: (a) To what extent does the SPELL early language and preliteracy intervention increase children's language and preliteracy skills relative to business-as-usual child care in Denmark? (b) To what extent does additional PD for educators or inclusion of a home-based companion program enhance the effects of SPELL on children's outcomes? (c) To what extent do child-level risk factors moderate SPELL impacts for children, specifically children's SES and language status? (d) To what extent is children's intervention exposure to SPELL associated with language and preliteracy outcomes for children?

Method

Participants

The multisite, multicohort study was designed using a priori power analysis with a target of 80% power to detect an effect size of $d = .20$ or greater in outcome variables. This resulted in the enrollment of 144 childcares serving children 3–6 years of age, with random assignment at the level of the center; this resulted in 36 childcares per each of four conditions. The centers, serving a total of 7,120 children, were recruited over two consecutive cohorts to distribute activities over time. Data were collected between November 2012 and June 2013. The number of childcares per condition varied slightly at pretest (business-as-usual [BAU], $n = 36$; SPELL, $n = 38$; SPELL + PD, $n = 34$; SPELL + HOME, $n = 34$). Data for all randomized childcares, educators, and children for whom background data could be obtained ($n = 7{,}076$) appear in Table 1.

Attrition Following Random Assignment

As expected in large-scale trials, this study had some missing data. First, some data were missing due to attrition of childcares, with eight leaving after randomization occurred (BAU, $n = 1$; SPELL, $n = 2$; SPELL + PD, $n = 3$; SPELL + HOME, $n = 2$): six left due to leadership or staff transitions and difficulties, one left for no provided reason, and one provided no posttest data (although they reportedly were obtained). Second, pretests and posttests were missing for some children, due to illnesses or other absences at each time point; overall, pretest was available for 6,483 children (92% of recruited sample), and pretest and posttest were available for 5,359 children (76% of recruited sample; see Tables S1–S4, for attrition data and analyses).

We evaluated differential attrition rates from randomization to posttest following the guidelines of the U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). for each of the three intervention conditions compared to the BAU. These guidelines differentiate between low attrition, which are attrition rates that are unlikely to result in bias, and high attrition; high attrition is when attrition levels are likely to result in bias. These levels are calculated

Table 1
Baseline Characteristics of Recruited Children, Child Cares, Classrooms, Child Care Educators, and Groups in Four Conditions

	BAU	SPELL	SPELL + PD	SPELL + HOME
Child cares, *n*[a]	36	38	34	34
Children, *n*	1,850	1,943	1,696	1,587
Classrooms per childcare, *M*	2.6	2.7	2.7	2.6
Educators per classroom, *M*	–	1.9	1.8	1.8
Groups per educator, *M*	–	1.8	1.9	1.8
Children per group, *M*	–	5.4	5.4	5.4
Educator characteristics				
% > 10 years teaching experience	62.1	58.5	54.3	54.8
% BA or higher education	78.3	74.3	77.0	66.7
Child characteristics				
% Boy	55.3	50.1	52.4	52.1
Age in months (*SD*)	53.8 (10.5)	52,6 (10.8)	53.9 (10.5)	52.9 (10.8)
Maternal education[b]				
% Low	15.5	14.7	14.7	17.5
% Low–mid	37.0	36.2	37.4	37.3
% High–mid	30.9	32.4	34.2	32.0
% High	16.6	16.8	13.6	13.2
Family income as mean quintile (*SD*)	2.9 (1.4)	2.9 (1.4)	2.9 (1.3)	2.8 (1.4)
% DLL background	10.7	11.5	12.9	12.1
Range of % DLL per child care	1–92	1–87	1–78	2–88

Note. Group characteristics were not significantly different (uncorrected *p*s > .2). BAU = business-as-usual; DLL = dual language learners; PD = professional development. [a]The unevenness was caused by (a) the formal merger around intervention start of two neighboring child cares (SPELL and SPELL + HOME), who therefore both did the SPELL condition, (b) one administrative error in the substitution of a child care in SPELL + PD by which the substitute child care did the SPELL condition, and (c) the dropout of two childcares (SPELL + PD, SPELL + HOME) following randomization and just prior to the start of intervention. [b]Compare education percentages for Denmark's population of women age 25–49 years: low (primary school), 20.0%; low–mid (high school, vocational education), 39.0%; high–mid (e.g., professional BA such as teacher), 25.7%; high (BA and advanced university education), 15.2%.

using both liberal and conservative approaches. We calculated an overall mean attrition rate for each pairwise comparison (e.g., SPELL + HOME vs. BAU) and interpreted them relative to the WWC standards. From recruitment to posttest, all three attrition rate comparisons were low at the child care level, even under conservative boundary assumptions (see Table S1). At the child level, differential attrition rates were low under liberal assumptions for SPELL and SPELL + PD, and low under conservative assumptions for SPELL + HOME. From pretest to posttest, differential attrition was low under conservative assumptions for all comparisons at both the child care and the child levels (see Table S2). We compared attrition rates across the four study conditions in terms of gender, age, DLL status, maternal educational background (low, high), and household income (low, high). There was no differential attrition across conditions for any background variable (all *p*s > .05), except for household income (see Table S3). For this variable, slightly higher levels of attrition ($p = .05$) were seen in the BAU condition from recruitment to posttest. Finally, comparisons of pretest scores for children included in the study, representing those with both pretest and posttest scores, and children who attrited following pretest, representing those with only a pre-test score, showed only negligible differences in baseline performance across the main study outcomes and the four study conditions (see Table S4).

Final Participant Pool for Analyses

The main research questions were evaluated for all children who had both pretest and posttest scores on a given subscale; this ranged from 1,742 for the sound discrimination measure to 5,283 for the vocabulary measure (see Table 2). The lower numbers of children for four of the seven measures were not due to attrition but rather administration of these measures to children based on age, in that older children received different measures than younger children. Other missing data were due to child absences during days of testing, as well as attrition from the study. Although it is common for

Table 2
Pretest and Posttest Scores on Study Measures by Condition

		BAU		SPELL		SPELL + PD		SPELL + HOME	
Measure	*n*	Pre *M* (*SD*)	Post *M* (*SD*)	Pre *M* (*SD*)	Post *M* (*SD*)	Pre *M* (*SD*)	Post *M* (*SD*)	Pre *M* (*SD*)	Post *M* (*SD*)
Preliteracy composite	5,306	−0.02 (0.86)	0.39 (0.83)	−0.06 (0.87)	0.54 (0.82)	−0.02 (0.88)	0.58 (0.79)	−0.10 (0.85)	0.49 (0.82)
Sound discrimination[a]	1,742	14.9 (2.1)	15.5 (1.3)	14.6 (2.3)	15.3 (1.9)	14.8 (2.1)	15.5 (1.3)	14.6 (2.0)	15.4 (1.4)
Rhyme detection[b]	2,923	10.4 (4.0)	11.6 (3.9)	10.0 (4.0)	12.5 (3.5)	10.4 (4.1)	12.8 (3.5)	10.5 (4.0)	12.4 (3.7)
Deletion[b]	3,432	4.2 (5.8)	6.9 (6.8)	3.6 (5.3)	7.7 (6.4)	4.1 (5.5)	8.1 (6.4)	3.1 (4.8)	6.8 (6.4)
Letter identification[b]	3,215	7.0 (3.6)	8.2 (3.5)	6.7 (3.6)	8.7 (3.3)	7.1 (3.7)	8.8 (3.3)	6.7 (3.7)	8.6 (3.5)
Language composite	5,315	−0.04 (0.84)	0.49 (0.86)	−0.17 (0.83)	0.48 (0.86)	−0.10 (0.84)	0.50 (0.85)	−0.18 (0.80)	0.39 (0.83)
Vocabulary[c]	5,283	43.6 (19.6)	52.8 (17.5)	40.6 (20.0)	51.3 (18.6)	42.0 (20.3)	52.2 (18.2)	40.2 (19.4)	50.2 (18.4)
Comprehension[c]	5,243	17.8 (4.2)	19.4 (3.7)	17.2 (4.4)	19.2 (4.0)	17.5 (4.3)	19.4 (3.8)	17.2 (4.4)	18.9 (4.0)
Communication[c]	5,049	29.5 (6.4)	31.3 (6.7)	28.3 (6.1)	31.0 (6.2)	29.1 (6.4)	31.4 (5.9)	28.6 (6.4)	30.7 (6.0)

Note. Data represent children with both pretest and posttest scores. BAU = business-as-usual; PD = professional development. [a]Administered only to 3-year-olds. [b]Administered only to 4- to 6-year-olds. [c]Administered to all children.

experimental trials to rely on imputation to address missingness, the current study design involved a data structure involving five levels, with analyses incorporating all five levels. As standard software for multiple imputation does not allow a five-level nested design, we included in each of our analyses all children for whom information on all variables involved in the analysis were available. Given that missingness based on attrition did not meet levels of concern with respect to bias, and the overall sample size should provide sufficient power to assess the primary research aims, our inability to impute data presents a small limitation to the study.

Procedures

The primary procedures of interest involved implementing the four study conditions and monitoring treatment implementation. The procedures are described below.

SPELL

Children whose child care was assigned to one the three variations of the SPELL intervention participated in a 20-week intervention featuring an explicit scope and sequence of language and preliteracy instruction during forty 30-min lessons implemented twice weekly by their educator. Children received the lessons in small groups, with each child assigned to one of four groups per classroom; with two educators per classroom, each was responsible for two groups. As noted previously, the SPELL lessons were provided to teachers on an iPad with an app featuring (a) lesson plans and (b) logging tool to complete implementation notes following each lesson.

Regarding the lessons, these involved a soft-scripted sequence of step-by-step instruction featuring a before, during, and after reading activity as well as suggested language that educators could use to support children's learning during each activity. In addition, a "Learners' Ladder" accompanied each lesson to provide strategies to educators by which to differentiate instruction. Regarding the logging tool, educators maintained obligatory implementation notes on the iPad that recorded child attendance, perceived child engagement during the lesson, and their usage of specific scaffolding strategies toward individual children. Educators could not advance to a new lesson without completing the implementation notes. Educators also tracked children's individual progress toward each SPELL objective within each of the four learning domains addressed in SPELL on the iPad three times during the 20-week intervention period, after Lessons 6, 20, and 36, using an informal assessment indication of whether the child "never," "sometimes," or "often" demonstrated the skill. Both implementation notes and progress notes were automatically wired to a database.

To facilitate educators' implementation of SPELL, prior to the start of the project each received all of the intervention materials and participated in a 2-day 14-hr training workshop. The workshop introduced educators' to the four SPELL learning domains and objectives, how to follow a lesson format, and how to use the "Learners' Ladder" to differentially support children. The workshop also provided educators with training in the app.

SPELL + PD

Educators in SPELL + PD implemented all SPELL components but received two additional days of intensive training in intervention Weeks 3 and 6, corresponding to about 14 extra hours of support. The additional training was designed to increase the quality of the intervention delivery by focusing on enhancing differentiation of instruction using the "Learners' Ladder" scaffolding strategies, especially for at-risk children. Educators were provided with a language profile of each child in their groups and received further training in practicing differentiation strategies with these children; training involved role play, observation of others, and observing themselves via videos collected in their classrooms. In addition, educators partook in ongoing brief sessions with other SPELL + PD participants, working in pairs, to conduct assignments related to SPELL implementation.

SPELL + HOME

In SPELL + HOME, educators implemented all SPELL components. In addition, parents implemented SPELL activities twice weekly at home to complement the SPELL intervention that their children received in their classrooms. Parents received companion materials, including the same 10 books and 20 lesson plans (translated into seven languages) that addressed the same learning domains and learning objectives as in SPELL. Their children's educators were responsible for provision of these materials and collection of ongoing logs from parents regarding use of these materials. To support their implementation, they received a SPELL calendar with stickers to mark each session. Besides access to model videos, the parents were not offered any specific training. The home activities were aligned in time with the SPELL lessons in the child cares (see also Bleses et al., 2014). Parent implementation was measured via a weekly paper log with a checklist of required activities; however, parent provision of the log back to educators was very low and/or educators did not provide parental logged data to the researchers; therefore, there was a great deal of missing data, and parental implementation was not assessed quantitatively.

Intervention Fidelity and Child Exposure

Intervention fidelity was documented in terms of the number of completed lessons; this was obtained from the educators' obligatory completion of implementation notes after each lesson on the iPad. Of the 40 possible lessons, educators in SPELL, SPELL + PD, and SPELL + HOME completed 75% (M = 30.1, SD = 10.1), 69% (M = 27.6, SD = 9.0), and 83% (M = 33.3, SD = 90.5) of lessons, respectively. The 10% of educators with the lowest implementation fidelity completed 15, 18, and 13 lessons in each condition, respectively, and the 10% of educators with the highest implementation fidelity in all three conditions completed all 40 lessons.

In addition, educators' implementation of specific components of lessons, including use of SPELL materials and attention to lesson objectives, was also derived from the implementation notes. In SPELL, SPELL + PD, and SPELL + HOME, educators reported implementing all components of a lesson in 76% (SD = 28), 71% (SD = 27), and 75% (SD = 26) of the lessons, respectively. There was no appreciable difference among conditions. Additionally, research staff observed and rated adherence to the intervention on the basis of a checklist developed for this specific purpose via video recordings of lessons submitted by the educators at the beginning, middle, and end of the 20-week program. The checklists examined teachers' adherence to the step-by-step elements of the lesson plans as well as explicit attention to the two instructional targets per lesson. The reliability of the checklist coding was determined by double coding a randomly selected 10% of video recordings; agreement between two individual coders ranged from 77% to 100%. Based on observational data on 234 lessons, educators obtained adherence ratings of 94% (SPELL) and 99% (SPELL + PD, SPELL + HOME) in terms of how well they followed the elements in the lesson plan, whereas the extent to which educators addressed the learning objectives was somewhat lower (65% in SPELL, 75% in SPELL + PD, and 60% in SPELL + HOME). There were no significant differences across the three SPELL conditions.

Children's exposure to the intervention within their classrooms, based on attendance during lesson implementation, was also assessed. On average, children in SPELL, SPELL + PD, and SPELL + HOME were exposed to 64% (M = 25.6, SD = 9.48), 58% (M = 22.6, SD = 9.28), and 70% (M = 27.6, SD = 9.26) of lessons, respectively. The 10% children with the lowest exposure received 12, 10, and 13 of lessons, respectively, and the 10% children with the highest exposure received 36, 35, and 37 of lessons, respectively. The exposure of children to the intervention did not significantly differ among conditions.

Measures

Measures used in this study were of two types: measures of children's language and preliteracy skills, and background information for children and parents.

Children's language and preliteracy skills were assessed pre- and postintervention by their educators using a published assessment instrument, Language Assessment of Children (LA; Bleses, Vach, Jørgensen, & Worm, 2010). The LA is already administered by educators in the vast majority of Danish municipalities as part of a national screening program in child cares, including all but one municipality included in the present SPELL trial. Educators in this municipality therefore received instruction in how to use the LA. Two age-dependent versions of the LA were administered. Four subscales were administered to 3-year-old children: sound discrimination (of initial phonemes in words, maximum score 16), vocabulary (expressive, maximum score 40), comprehension (of words and complex sentences, maximum score 12), and communication (questionnaire with communicative strategies, maximum score 40). Six subscales were administered to 4- to 6-year-old-children: rhyme detection (maximum score 17), deletion (of words, syllables and sounds, maximum score 20), letter identification (maximum score 12), vocabulary (expressive, maximum score 76), comprehension (of words and complex sentences, maximum score 27), and communication (questionnaire with communicative strategies, maximum score 76). With the exception of communication, children were directly assessed via picture identification or picture elicitation tasks.

For the present purposes, two composite measures were derived from the seven subscales: a preliteracy composite based on averaging standardized scores for sound discrimination, rhyme detection, deletion, and letter identification, and a language composite based on averaging standardized scores for vocabulary, comprehension, and communication. Standardized scores were based on age- and gender-specific norms based on pretest. Internal consistency coefficients (Cronbach's alpha) for subscales were between .75 and .91, and correlations between subscales were between .25 and .70. The concurrent correlations of the language subscales with the Peabody Picture Vocabulary Test, 4th ed. (PPVT–4; Dunn & Dunn, 2012) were .55 for vocabulary and .57 for comprehension; correlations with the Expressive Vocabulary Test, 2nd ed. (EVT–2; Williams, 2007) were .42 for vocabulary and .39 for comprehension. For the preliteracy subscales, the correlations with PPVT–4 were .33 for rhyme detection, .49 for deletion, and .39 for letter identification; the correlations with EVT–2 were .18 for rhyme detection, .29 for deletion, and .33 letter identification (see Tables S5–S10).

Background information for children and parents with respect to SES and language status (DLL or not) was also collected. The former was indexed based on maternal education (see Table 1); for the latter, parental immigration status served as the index. This information was obtained from Statistics Denmark, using the Danish Central Personal Number System.

Analytic Strategy

The data collected in this effectiveness trial were nested in a complex way, with children nested in small groups in which intervention was received, the educators who delivered the intervention, their classrooms, and their centers. To address this nested data structure, the research questions were addressed using five-level hierarchical linear models (HLM; Raudenbush & Bryk, 2002), with child represented as Level 1, group as Level 2, educator as Level 3, classroom as Level 4, and child care center as Level 5 in the analyses. Levels 4 (classroom) and 5 (center) variables were treated as random effects for all three conditions. Levels 2 (group) and 3 (educator) variables were treated as random effects for the two intervention conditions only, as children in the BAU condition were not assigned to groups or to educators. Intraclass correlations (ICCs) provide evidence for variance at each of the five levels and the extent to which the analyses should consider variables at each level. At the center level, ICCs ranged from 0 on sound discrimination to .11 on rhyme detection; at the classroom level, ICCs ranged from 0 on sound discrimination to .16 on communication; at the educator level, ICCs ranged from .0 on letter identification to .16 on sound discrimination; at the group level, ICCs ranged from .01 on sound discrimination to .07 on vocabulary; and at the child level, ICCs ranged from .63 on communication to .86 on letter identification. These values indicate that the majority of variance was between children, but that each level contributed some true variation.

Covariates were included in analyses based on the specific research question being addressed. Specifically, for all four questions, the pretest value corresponding to the posttest value of interest was always included as a covariate. When the

comprehension measure was the outcome variable of interest, the version of LA was included as a covariate to take into account the different number of items in the two age-dependent versions. When maternal education was entered into a model as a covariate, the child's DLL status was included as an additional category, such that only the effect of maternal education for Danish children was estimated. This was necessary because there was a lack of information about maternal education for many DLL children in Statistics Denmark (which was the source for these particular data points); this is because Statistics Denmark has data on maternal education only for those mothers who complete their schooling in the country. Put differently, for the many DLL children, we did not have information about maternal education.

Results are interpreted in effect-size units, based on Cohen's *d*, based on dividing the HLM effect estimates by the standard deviation of the individual pretest (Cohen, 1988). Pairwise comparison *p* values were based on the Kramer–Tukey method for correcting for multiplicity (see Jaccard, Becker, & Wood, 1984). All analyses were conducted using the STATA mixed command except that Kramer–Tukey *p* values were calculated using SAS PROC MIXED, AS Institute Inc., Cary, NC, USA.

Results

Preliminary Analyses

Preliminary analyses were conducted to determine the extent to which educators, classrooms, and children were initially equivalent across the four study conditions. Table 1 presents descriptive statistics for those variables investigated in preliminary analyses. HLM analyses and logistic regression with robust standard errors indicated that there were no significant differences among the four conditions on any continuous or binary characteristic shown in Table 1 (all *p*s > .20). Similarly, an analysis of pretest values showed no significant differences (all *p*s > .24). In these preliminary analyses, it was noted that some of the subscales exhibited ceiling effects already at pretest. For 3-year-olds, the median score on sound discrimination was 16, equal to the maximum score on the scale; children at the 25th percentile obtained 14 items correct. The comprehension measure also exhibited ceiling effects for 3-year-olds (median = 16, maximum = 20). No scale exhibited ceiling effect for 4-year-olds, but for 5-year-olds, there were ceiling effects for letter knowledge (score on 75th percentile = 11, maximum = 12), rhyme detection (score on 90th percentile = 16, maximum = 17), and a less pronounced ceiling effect for vocabulary (score on 90th percentile = 69, maximum = 76). Given these ceiling effects on the pretests, a proportion of the children (10%–75% depending on subscale) had no possibility of showing improvement on posttests. However, we maintained these measures in our analyses as they had been proposed in the trial registry (Bleses et al., 2014). It is worth noting that the effect estimates therefore are conservative.

Main Effects of SPELL

To address the first and second research questions, we examined the extent to which three variations of SPELL had positive effects on children's outcome as compared to BAU. Table 2 provides pre- and posttest data for the seven subscales of the LA and the two composites. We estimated nine separate HLMs, one for each outcome variable and for each of the two composites, with the following fixed effect covariates included at each level: at Level 1 (child level), the corresponding pretest score, age, version of LA (depending on age; only relevant for comprehension subscale), and gender; and at Level 5 (child care level), the condition to which the child care was assigned, the cohort in which the child care participated, and the municipality of the child care. Results for the nine HLM analyses appear in Table 3.

Looking first at the contrasts specific to SPELL and BAU, the results show that children who experienced SPELL outperformed those receiving BAU on three subscales (rhyme detection, $d = .20$; deletion, $d = .27$; letter knowledge, $d = .19$) in addition to the preliteracy composite ($d = .22$). Children in SPELL also scored higher than those in BAU on all other indices, with the exception of sound discrimination ($d = -.05$ to .07), with effect-size contrasts ranging from .08 (communication) to .10 (comprehension), although the contrasts were not statistically significant. For the SPELL + PD and BAU contrasts, results show that children in the SPELL + PD condition outperformed those receiving BAU on the rhyme detection ($d = .25$), deletion ($d = .31$), and letter knowledge ($d = .17$) subscales and on the preliteracy composite ($d = .27$). For the SPELL + HOME and BAU contrasts, children in the former condition outperformed those in the latter on the deletion ($d = .16$) and letter knowledge ($d = .21$) subscales and on the preliteracy composite ($d = .21$); no other comparisons were significant.

Table 3
Effect-Size Estimates for Three SPELL Conditions Compared to Business-as-Usual (BAU)

Variable (BAU *n*)	SPELL	SPELL + PD	SPELL + HOME
Preliteracy composite (*n* = 1,324)	.22 (1,533)**	.27 (1,317)**	.21 (1,132)**
Sound discrimination (*n* = 407)	−.05 (516)	.07 (441)	.01 (378)
Rhyme (*n* = 722)	.20 (850)*	.25 (749)**	.11 (602)
Deletion (*n* = 867)	.27 (996)**	.31 (839)**	.16 (735)*
Letter identification (*n* = 812)	.19 (959)**	.17 (777)*	.21 (667)**
Language composite (*n* = 1,329)	.16 (1,539)	.13 (1,317)	.04 (1,130)
Vocabulary (*n* = 1,325)	.09 (1,535)	.09 (1,315)	.05 (1,108)
Comprehension (*n* = 1,310)	.10 (1,523)	.08 (1,304)	.04 (1,106)
Communication (*n* = 1,250)	.08 (1,511)	.05 (1,270)	.01 (1,018)

Note. PD = professional development. *$p < .05$. **$p < .01$.

With the exception of sound discrimination, which generally showed no effect, there was not a clear relationship between the existence of ceiling effects and the effect size for measures.

When comparing the three SPELL conditions to one another using pairwise comparisons, only one reached statistical significance, namely the sound discrimination measure. For this comparison, the SPELL + PD condition was favored over the SPELL condition (*d*s = .07 and −.04, respectively, $p = .044$). All other comparisons were not statistically significant. Therefore, although the SPELL + PD condition had the highest effect sizes relative to BAU, it did not distinguish itself from the other two SPELL conditions with respect to impacts on children's language and preliteracy skills. Thus, findings suggest that the SPELL + HOME and SPELL + PD supplements did not have practical relevance for enhancing effects across the whole sample over SPELL itself.

Effect Moderation by Child Characteristics

The third research question sought to determine whether non-at-risk and at-risk children (indexed here as low maternal education) and DLL children would benefit from the intervention equally well. To address this question, two HLMs, one per each of the two composite outcomes (preliteracy composite, language composite), were estimated in which child-level covariates of DLL status and maternal education were included both as a main effect and as interactions with condition. The continuous variables were grand mean centered. For the sake of parsimony and power for analysis of interactions, we collapsed all three SPELL conditions into a single intervention condition for these analyses, given that the three interventions did not distinguish themselves from one another with respect to child outcomes. Additional covariates included in the models included children's pretest scores, age, gender, and intervention.

Table 4 provides the unstandardized coefficients (*B*), standard errors, and *p* values for each of the predictor variables and interaction terms. Note that in these models, the estimated effect of SPELL is the difference between the BAU condition and the pooled intervention group for a non-DLL boy with average pretest score, and average maternal education. For the interaction terms, the interactions including pretest, age, gender, DLL status, and maternal education describe how the slopes of

Table 4
Hierarchical Linear Model Results for Predicting Preliteracy Composite and Language Composite From Child-Level Characteristics

	Preliteracy composite		Language composite	
	B (*SE*)	*p*	*B* (*SE*)	*p*
Base model				
Pretest	.50 (.02)	< .001	.72 (.02)	< .001
Age	.03 (.02)	.201	−.08 (.02)	< .001
Gender	−.05 (.04)	.138	−.04 (.03)	.108
DLL status	−.10 (.08)	.188	−.24 (.06)	< .001
Maternal education[a]	.09 (.02)	< .001	.06 (.02)	.001
Interaction model				
Intervention (SPELL)	.19 (.05)	< .001	.06 (.05)	.224
SPELL × Pretest	−.05 (.03)	.044	.00 (.00)	.929
SPELL × Age	.00 (.00)	.466	−.00 (.00)	.748
SPELL × Gender	−.00 (.04)	.925	−.05 (.03)	.118
SPELL × DLL	−.16 (.09)	.066	−.05 (.07)	.505
SPELL × Maternal education[a]	−.02 (.03)	.361	−.01 (.02)	.649

Note. DLL = dual language learners. [a]Maternal education describes the effect in non-DLL children only, as maternal education variable was not available for DLL children.

these variables change if we move from BAU to the intervention conditions. For the interaction including the DLL variable, the interaction can be best interpreted as the change of the SPELL effect when moving from a nonimmigrant to an immigrant child.

Among the variables included in the first model, for which the preliteracy composite served as the outcome variable, children's pretest scores ($p < .001$) and maternal education ($p < .001$) served as unique, significant predictors of the outcome variable; in addition, intervention served as a unique, significant predictor of the outcome ($p < .001$) as well as pretest scores ($p = .044$). Figure S1 depicts the nature of the interaction between pretest scores and the intervention effect, which was the only significant interaction, showing that children with lower pretest scores gained more from the intervention than children with higher pretest scores. However, this interaction may be driven by the ceiling effect seen for some measures included in the preliteracy composite, thus interpretation of the result should be tempered.

Variables included in the second model, with language composite as the outcome, included pretest ($p < .001$), age ($p < .001$), DLL status ($p < .001$), and maternal education ($p = .001$) as independent, significant predictors of the outcome variable. Neither intervention ($p = .224$) nor any of the other variables interacted significantly with the outcome. In sum, neither maternal education level (SES) nor DLL background significantly moderated the SPELL treatment effects.

Relations Between Intervention Exposure and Child Outcomes

The fourth research question considered the role of children's exposure to the intervention, namely the number of lessons in which the child participated, as a potential variable of influence on children's outcomes. To examine the relations between children's intervention exposure and intervention outcomes, two additional HLMs were estimated, one for each of the posttest composites (preliteracy composite, language composite). For these models, only children within the three SPELL conditions were included ($n = 3{,}982$ for the preliteracy composite, $n = 3{,}986$ for the language composite). Predictors included in the base model were the corresponding pretest, age, gender, DLL status, and maternal education as well as the interaction terms for Intervention Exposure (number of lessons completed at the child level) × DLL status and intervention exposure × maternal education. The interactions allowed us to consider whether increased exposure to SPELL for either all or a subsample of children was associated with improved outcomes. The interaction terms describe how this number changes if we consider a non-DLL versus DLL child (intervention exposure × DLL Status) or if we increase maternal education by one level (intervention exposure × maternal education); that is, if the effect of increased exposure varies depending on immigrant status or maternal education.

Results, as presented in Table 5, showed that among base model variables included in the first model, for which the preliteracy composite served as the outcome, children's pretest scores ($p < .001$), gender ($p = .006$), DLL status ($p < .001$), maternal education ($p < .001$), and intervention exposure ($p < .001$) served as a unique, significant predictors of the outcome variable. Neither DLL status ($p = .882$) nor maternal education ($p = .403$) interacted significantly with intervention exposure to predict the outcome variable.

Among base model variables included in the second model, for which the language composite served as the outcome, children's pretest scores ($p < .001$), age ($p < .001$), DLL status ($p < .001$), maternal education ($p < .001$), and intervention exposure ($p < .001$) served as a unique, significant predictor of the outcome variable. DLL status interacted significantly with intervention exposure ($p = .04$) to predict the outcome, whereas maternal education did not ($p = .66$). Figure S2 illustrates the larger effect of exposure for DLL children's language composite scores than for non-DLL children. Intervention

Table 5
Hierarchical Linear Model Results for Predicting Children's SPELL Outcomes From Intervention Exposure

	Preliteracy outcomes		Language outcomes	
	B (*SE*)	*p*	*B* (*SE*)	*p*
Pretest	4.5 (0.1)	< .001	7.2 (0.1)	< .001
Age	0.0 (1.2)	.118	−0.1 (0.0)	< .001
Gender	−0.5 (0.2)	.006	−0.0 (0.1)	.744
DLL status	−2.4 (0.4)	< .001	−2.6 (0.3)	< .001
Maternal education[a]	0.7 (0.1)	< .001	0.4 (0.1)	< .001
Intervention exposure	0.1 (0.0)	< .001	0.1 (0.0)	< .001
Intervention Exposure × DLL Status	0.0 (0.0)	.882	0.1 (0.0)	.042
Intervention Exposure × Maternal Education	−0.0 (0.0)	.403	0.0 (0.0)	.665

Note. DLL = dual language learners. [a]Maternal education describes the effect in DLL children only.

exposure accounted for 9% of the variation in outcomes in rhyme detection, 14% in deletion, 17% in letter identification, and 19% in vocabulary. The strength of the intervention was consistently associated with children's exposure. We found significant effects on language composite scores for children who were exposed to more than 20 lessons corresponding to 10 hr of intervention (SPELL condition [d = .24] and the SPELL + PD condition [d = .19] for the 21 + subsample), which further suggests that had the dosage of intervention been higher, we might have seen larger, overall effects for the language outcomes. Correlational analyses indicate that child background characteristics predicted amount of exposure; in particular, DLL status showed that DLL children were exposed to fewer lessons than non-DLL children (r = −.15), while at the same time the effect of exposure was stronger for DLL children than other children. It should be acknowledged that all results for exposure are correlational not experimentally manipulated.

Discussion

The present study is distinctive, in that it was designed to investigate the extent to which language and preliteracy interventions positively affect young children when taken to scale under real-world circumstances with an unselected sample in a European setting. This investigation was conducted throughout the country of Denmark, which distinguishes it from prior efficacy and effectiveness trials largely situated within the United States. Several contrasts to the United States' context warrant reference in understanding the contribution of this work. First, Danish childcares are publicly funded and available at scale; to this end, nearly all preschool-aged children attend, thus allowing us to examine effects for a general population of youngsters instead of children attending programs with targeted enrollment, such as Head Start in the United States and many state-supported preschool programs. Second, these childcares are situated in mostly residential areas across Denmark, thus their socioeconomic composition is solely determined by the composition of the neighborhood. Thus, Danish childcares typically serve a socioeconomically diverse group of children, even though childcares with a high proportion of at-risk children also exist. In such a context, we were able to investigate intervention effects for all children within these child care settings while also exploring whether certain groups of children, as a function of family and child characteristics, may derive particular benefit. Finally, the explicit language and preliteracy instructional approach utilized in the SPELL intervention is not one typically utilized in the Danish context, in which early educators tend to use child-oriented pedagogies and emphasize social development with little to no attention to academically oriented outcomes. The trial results are compelling in this regard, as it helps us to understand whether interventions developed in one context can be applied to a novel and distinct context.

The principal finding of this effectiveness study is that SPELL had significant impacts on two fundamental preliteracy skills, phonological awareness and letter knowledge, both of which were explicitly targeted by the intervention. Children in SPELL classrooms outscored children in BAU classrooms by about 0.25 *SD* unit, on average, for these measures. Although children in the three SPELL conditions also scored higher (d = .05–.10) than those in the BAU on vocabulary, also targeted by the intervention, none of these contrasts reached significance. These effects are generally comparable to results recently reported for a larger scale effectiveness trial for RIA (the intervention from which SPELL was adapted) involving preschoolers from low-SES backgrounds (n = 406) in Appalachian communities (Mashburn, Justice, McGinty, & Slocum, 2016). Effect-size estimates indicated that exposure to RIA improved children's preliteracy skills on measures of alphabet knowledge and print concepts on the magnitude of .18 and .25 *SD* units, respectively, as compared to a control. Similar to the present study, no positive effects were seen on measures of early language skill. However, Danish children exposed to SPELL showed significant improvements in phonological awareness relative to controls, but no such effect was found for American children exposed to RIA. It may be that SPELL was particularly beneficial to children's development of phonological awareness because it, in contrast to many U.S. preschools, is seldom taught within BAU classrooms in Denmark.

Although positive effects on preliteracy skills were observed in the present trial for children exposed to SPELL, the effect sizes observed were considerably lower than those obtained in prior efficacy trials, which range from .25 for language outcomes to .87 for phonological awareness (see NELP, 2008). Similarly sized effects were seen in an early efficacy study involving RIA in rural communities (e.g., .35 for grammar, .44 for print concepts; see Justice et al., 2010). The present findings confirm, as would be expected, that taking systematic explicit language

and preliteracy interventions to scale is likely to decrease the size of the effects observed but also that favorable effects are still observed in real-world contexts. Indeed, benchmarking the effects against the randomized controlled effectiveness trials described in the Appendix, the effect sizes are quite comparable, even though the SPELL intervention had a duration of just 20 weeks versus the school year duration for the other effectiveness studies.

An additional and important result derived from this study, and one that stands in contrast to previous research, is that neither SES indexed as maternal education nor DLL background significantly moderated the intervention effect as pooled across the three intervention groups. Unlike many prior studies, the current effectiveness trial was sufficiently powered to assess moderating variables. Therefore, the lack of moderation of intervention effects by children's SES or language status is an important contribution to the literature, as it suggests that low-SES children and DLLs benefited to the same extent as other, more advantaged children. Stated conversely, the advantaged children in the sample benefited similarly to the more at-risk participants. Of course, we must caution that such effects may not generalize outside of the Danish context, as children who are considered low SES and DLLs in other contexts may differ in important ways from Danish children who are low SES and/or are DLLs. Nonetheless, given that intervention effects were not contingent on key child-level characteristics, this adds to the practical relevance of the effects. As some have argued, a phenomenon with a small-to-modest effect on many individuals may have as large an impact collectively as a phenomenon with a large effect on only a few individuals (Early Child Care Research Network, 2006).

A third finding of interest in this effectiveness study concerns issues related to implementation. Across the three groups of educators using the experimental intervention, low levels of implementation fidelity were observed. This was not unexpected for several reasons. First, from a design perspective, effectiveness studies are designed in such a way so as to reduce the researchers' level of control over implementation of the intervention. With relatively low levels of researcher involvement in and control over implementation, the primary goal of an effectiveness trial is to determine whether effects seen in tightly controlled investigations are maintained. Second, from a contextual perspective, Danish early educators are not accustomed to using explicit, systematic interventions, and implementation of structured curricula in these settings is atypical. Therefore, the educators may be unaccustomed to use explicit and systematic practices. The organizational structure in many Danish child cares is not supportive of implementing structured interventions like SPELL, which utilize structured small-group activities. When adopting novel language- and preliteracy instructional interventions, it can take time for early educators to feel comfortable applying the approach and fidelity may be reduced as a result (Pence et al., 2008). More than this, however, the early educational context within Denmark emphasizes democratic values and social development, with academic foci treated warily. Therefore, some early educators are reluctant to engage in explicit instruction that is "school like." Thus, it is not completely surprising that children in this study were exposed to an average of only 25 of the 40 planned 30-min SPELL lessons, corresponding to about 12.5 hr of intervention in total. These relatively low levels of fidelity may not only attenuate effects of the intervention on targeted child outcomes but also represent the level of implementation that may be realistically expected in everyday circumstances.

To this point, children's actual exposure to the intervention was positively associated with their gains over time for both the language and preliteracy composites. Simply, children who experienced a larger percentage of SPELL lessons appeared to benefit more from the intervention, consistent with a dosage effect. To some extent, this effect may reflect a true treatment effect, with greater opportunities to learn associated with greater learning, as would be expected based on prior dosage work. For instance, studies of vocabulary development have shown improved learning of novel words with additional learning opportunities (Loftus-Rattan, Mitchell, & Coyne, 2016). On the other hand, it is also possible that this effect represents a key third variable, namely child care attendance. In early childhood settings in the United States, children's attendance can be highly variable; for instance, one study of Head Start participants reported a mean daily attendance rate of 85% (Hubbs-Tait et al., 2002). Importantly, attendance rates, even when controlling for key background factors, are associated with children's growth in social skills (Hubbs-Tait et al., 2002) and language skills (Logan, Piasta, Justice, Schatschneider, & Petrill, 2011). In the present study, we cannot disentangle the effects of regular attendance in childcare from SPELL exposure, and it may be possible that the two interact in important ways. Enhancing our understanding of intervention exposure for young children should be continually assessed in effectiveness studies.

It is worth considering what might be done in future work to enhance implementation by child-care educators, as one avenue for improving children's intervention exposure. First, improving educators' experience with digital technologies may result in higher levels of implementation of the app-based intervention. In general, early educators do not use many digital technologies in their classroom; however, those with high levels of technology-based experience have higher rates of usage (Blackwell, Lauricella, & Wartella, 2014). Second, exploring other avenues for explicitly teaching children early language and preliteracy skills may improve implementation; for instance, perhaps such explicit teaching can be provided during play or dramatic activities. Danish teachers are generally reluctant to engage in explicit instruction, which may result in low implementation rates.

Implementation science research can provide an important mechanism for exploring the extent to which these factors may present barriers to educators' implementation of novel language and preliteracy interventions. The Theoretical Domains Framework (see French et al., 2012) is one approach used within implementation science to identify determinants of behavior that negatively affect intervention implementation. The Theoretical Domains Framework can be used to identify both behavioral determinants that detract from implementation but also empirically informed approaches that work to affect these determinants. As early interventions are increasingly taken to scale outside of the context in which they were developed, efforts to understand and improve implementation processes should be a central focus (see Fixsen, Naoom, Blase, & Friedman, 2005).

Several additional findings also warrant discussion. In this study, we also tested the extent to which two additional components improved the effectiveness of SPELL. Neither of the two components resulted in significantly improved outcomes in children. Even though there were nominally higher effect sizes for phonological awareness subscales in SPELL + PD, the group differences reached significance in just a single case (SPELL vs. SPELL + PD for sound discrimination). Similarly to the results of Lonigan et al. (2011) and Assel et al. (2007), our study therefore found limited evidence for an effect of extended PD for educators to boost the effects of an intervention. Indeed, because the extra PD provided to educators did not seem to have value-added effects for the intervention, we can argue that the explicit and systematic instruction provided by SPELL was the most critical element of the intervention. Nonetheless, we do note that an inspection of effect sizes across all subgroups did show a consistent pattern of nominally higher effect sizes among children with educators in the SPELL + PD condition, at least for the preliteracy composite, perhaps indicating that providing educators with specific training in recognizing and practicing differentiation strategies may assist educators in meeting the needs of children. Given that children in the SPELL + PD condition were exposed to the fewest number of lessons, it is possible that the nominal advantage of the SPELL + PD group would have been larger and significant had they received the same number of lessons as the other two intervention groups. This hypothesis is further supported by the fact that even though the teachers in SPELL + PD completed fewer lessons, the quality of the completed lessons was somewhat higher, with teachers addressing learning objectives more frequently. Together, these results suggest that combining a cost-effective PD component to a systematic explicit curriculum, based on elements from a practice-based approach (Hamre et al., 2012; Landry et al., 2009, 2011), merits further investigation.

Turning to the SPELL + HOME condition, the effect sizes were generally no larger than seen, and even sometimes slightly smaller, in magnitude compared to the basic SPELL condition, similar to the findings of Manz et al. (2010). An inspection of effect sizes in subgroups of children did, however, revealed that in the SPELL + HOME condition, children from lower SES homes had nominally higher effect sizes than children from the high-SES homes; this may suggest that providing more learning opportunities to children via parallel activities in the home may benefit children from low-SES homes, consistent with recent work by Bierman et al. (2015). To this end, a high priority for future effectiveness studies of home educational components is to investigate both their overall effect and differential effects for subgroups while holding the center-based curriculum constant. Furthermore, the results of this study suggest exploring how parents might be engaged differentially depending on the needs of the children. There is some anecdotal evidence for this suggestion, as some educators commented that children whose parents appeared to deliver the SPELL + HOME intervention with high fidelity seemed to lose interest in SPELL during the course of intervention, perhaps from an "overdosage" of these learning opportunities.

Several limitations of this study warrant mention. First, as is typical of real-world studies,

attrition was an issue, although an analysis based on WWC guidelines suggests that these did not achieve a level so as to have biased the results. Second, as noted previously, some of the subscales exhibited ceiling effects at pretest, such that a percentage of children had no possibility of showing improvement on posttests. This may have led to an underestimation of effects. A related limitation is the lack of psychometrically well-established language and preliteracy measures to use with young, Danish children; thus, although the measures we did use have been piloted and show concurrent relations with some established measures, they have not been as rigorously developed as measures available in other international contexts. Third, the assessments administered to children were administered by their child care providers, and not trained research staff. Although we do not believe this biased results, as the educators were blind to the various conditions in the study and prior research shows strong congruence between educators' and researchers' assessments (Cabell, Justice, Zucker, & Kilday, 2009), confidence in these results would be boosted with independent assessments of children. Additionally, the absence of verifiable fidelity of the home component is also a limitation of the study. Finally, until long-term follow-up data are available, it is unknown whether the obtained effects will result in improved reading skills.

In sum, this effectiveness study was motivated by the need to enhance children's language and preliteracy skills during the preschool period, with a focus on those skills known to be foundational for later educational success. Although children from low-SES/DLL homes are at higher risk, lags in development can occur for children throughout the population, and improved educational attainment is also possible even for children doing relatively well. For these reasons, it is vital to develop effective interventions for language and preliteracy development which are of benefit for all children. The results of this study demonstrated that even when implemented at full scale in a heterogeneous sample, a semiscripted, skill-focused curriculum supplement requiring little training prior to implementation and no costly ongoing PD can be an economically reasonable means for improving preliteracy skills to the benefit of all children in child care. Although the intervention significantly impacted children's preliteracy skills, significant impacts were found for language skills only in children who were exposed to at least half of the 40 lessons. This highlights the importance of investigating factors in the local educational context that can support higher implementation fidelity of interventions such as SPELL. Despite the challenges of conducting rigorous effectiveness trials, this approach yields results that are highly relevant for policymakers, because these are designed to answer the most relevant questions for decision and budget makers. There is a great need for rigorous effectiveness studies with high level of both internal and external validity to test the generalizability of well-known curricula with established efficacy.

References

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing*, *20*, 463–494. https://doi.org/10.1007/s11145-006-9039-5

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, *9*, 43–61. https://doi.org/10.1207/s1532799xssr0901_4

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, *79*, 1802–1817. https://doi.org/10.1111/j.1467-8624.2008.01227.x

Bierman, K. L., Welsh, J. A., Heinrichs, B. S., Nix, R. L., & Mathis, E. T. (2015). Helping Head Start parents promote their children's kindergarten adjustment: The research-based developmentally informed parent program. *Child Development*, *86*, 1877–1891. https://doi.org/10.1111/cdev.12448

Blackwell, C. K., Lauricella, A. R., & Wartella, E. (2014). Factors influencing digital technology use in early childhood education. *Computers & Education*, *77*, 82–90. https://doi.org/10.1016/j.compedu.2014.04.013

Bleses, D., Dale, P. S., Højen, A., Justice, L. M., Jensen, P., & Andersen, T. L. (2017). Identifying effective instructional approaches in early language and literacy interventions: Meta-analysis of 25 years of research (1990-2014). Manuscript submitted for publication.

Bleses, D., Højen, A., Dybdal, L., Dale, P. S., Justice, L., Piasta, S., & Aktürk Ari, B. (2014). A randomized controlled trial of a large scale systematic and explicit language and literacy intervention in Danish daycares: The SPELL study. *British Journal of Education, Society & Behavioural Science*, *4*, 504–1524. https://doi.org/10.9734/BJESBS/2014/11691

Bleses, D., Højen, A., Jørgensen, R. N., Jensen, K. Ø., & Vach, W. (2010). *Sprogvurdering af 3-årige (09)—karakteristika og risikofaktorer [Language assessment of 3 year olds. Characteristics and risk factors]* (Vol. 10). Odense, Denmark: Center for Child Language.

Bleses, D., Vach, W., Jørgensen, R. N., & Worm, T. (2010). The internal validity and acceptability of the Danish

SI-3: A language screening instrument for 3-year-olds. *Journal of Speech, Language, and Hearing Research*, *53*, 490–507. https://doi.org/10.1044/1092-4388(2009/08-0132)

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development*, *85*, 1346–1356. https://doi.org/10.1111/cdev.12192

Buysse, V., Peisner-Feinberg, E., Páez, M., Hammer, C. S., & Knowles, M. (2014). Effects of early education programs and practices on the development and learning of dual language learners: A review of the literature. *Early Childhood Research Quarterly*, *29*, 765–785. https://doi.org/10.1016/j.ecresq.2013.08.004

Buysse, V., Sparkman, K. L., & Wesley, P. W. (2003). Communities of practice: Connecting what we know with what we do. *Exceptional Children*, *69*, 263–277. https://doi.org/10.1177/001440290306900301

Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools*, *40*, 161–173. https://doi.org/10.1044/0161-1461(2009/07-0099)

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*, 1407–1425. https://doi.org/10.1007/s11145-015-9576-x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Danish Ministry for Social Affairs. (2015). *Tal og statistik om dagtilbud [Numbers and statistics about daycares]*. Copenhagen: Danish Ministry for Social Affairs.

Dockrell, J. E., Stuart, M., & King, D. (2010). Supporting early oral language skills for English language learners in inner city preschool provision. *British Journal of Educational Psychology*, *80*, 497–515. https://doi.org/10.1348/000709910X493080

Dunn, L. M., & Dunn, D. (2012). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT–4)*. San Antonio, TX: Pearson.

Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD study of early child care and youth development. *American Psychological Association*, *61*, 99–116. https://doi.org/10.1037/0003-066X.61.2.99

Farver, J. A. M., Lonigan, C. J., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. *Child Development*, *80*, 703–719. https://doi.org/10.1111/j.1467-8624.2009.01292.x

Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language imparment: An experimental evaluation. *Journal of Speech, Language, and Hearing Research*, *36*, 141–157. https://doi.org/10.1044/jshr.3601.141

Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S., & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research*, *39*, 471–501. https://doi.org/10.1080/10862960701675333

Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, The National Implementation Research Network.

French, S. D., Green, S. E., O'Connor, D. A., McKenzie, J. E., Francis, J. J., Michie, S., . . . Grimshaw, J. M. (2012). Developing theory-informed behaviour change interventions to implement evidence into practice: A systematic approach using the Theoretical Domains Framework. *Implementation Science*, *7*, 1. https://doi.org/10.1186/1748-5908-7-38

Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N., & Carey, T. S. (2006). *Criteria for distinguishing effectiveness from efficacy trials in systematic reviews* (Vol. 12). Technical Reviews. (Prepared by the RTI-International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016.) AHRQ Publication No. 06-0046. Rockville, MD: Agency for Healthcare Research and Quality.

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. https://doi.org/10.3102/0002831211434596

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. https://doi.org/10.1037/a0027238

Hubbs-Tait, L., Culp, A. M., Huey, E., Culp, R., Starost, H. J., & Hare, C. (2002). Relation of Head Start attendance to children's cognitive and social outcomes: Moderation by family risk. *Early Childhood Research Quarterly*, *17*, 539–558. https://doi.org/10.1016/S0885-2006(02)00189-8

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88–110. https://doi.org/10.1080/19345740802539325

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise multiple comparison procedures: A review. *Psychological Bulletin*, *96*, 589. https://doi.org/10.1037/0033-2909.96.3.589

Justice, L. M., Kaderavek, J. N., Fan, X., Sofka, A., & Hunt, A. (2009). Accelerating preschoolers' early literacy development through classroom-based teacher–child storybook reading and explicit print referencing. *Language, Speech, and Hearing Services in Schools*, *40*, 67–85. https://doi.org/10.1044/0161-1461(2008/07-0098)

Justice, L. M., Logan, J. A., Kaderavek, J. N., & Dynia, J. M. (2015). Print-focused read-alouds in early childhood

special education programs. *Exceptional Children*, *81*, 292–311. https://doi.org/10.1177/0014402914563693

Justice, L. M., & McGinty, A. S. (2013). *Read It Again Pre-K! A preschool curricular supplement to promote language and literacy foundations*. Crane Center for Early Childhood Research and Policy. Retrieved from https://earlychildhood.ehe.osu.edu/files/2016/05/Curriculum-Supplement.pdf

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. https://doi.org/10.1044/0161-1461(2009/08-0058)

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, *101*, 448–465. https://doi.org/10.1037/a0013842

Landry, S. H., Swank, P. R., Anthony, J. L., & Assel, M. A. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Reading and Writing*, *24*, 971–1010. https://doi.org/10.1007/s11145-010-9243-1

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington, DC: National Center for Special Education Research.

Loftus-Rattan, S. M., Mitchell, A. M., & Coyne, M. D. (2016). Direct vocabulary instruction in preschool: A comparison of extended instruction, embedded instruction, and incidental exposure. *The Elementary School Journal*, *116*, 391–410. https://doi.org/10.1086/684828

Logan, J. A., Piasta, S. B., Justice, L. M., Schatschneider, C., & Petrill, S. (2011). Children's attendance rates and quality of teacher-child interactions in at-risk preschool classrooms: Contribution to children's expressive language growth. *Child & Youth Care Forum*, *40*, 457–477. https://doi.org/10.1007/s10566-011-9142-x

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. https://doi.org/10.1007/s11145-009-9214-6

Lonigan, C. J., & Phillips, B. M. (2016). Response to instruction in preschool: Results of two randomized studies with children at significant risk of reading difficulties. *Journal of Educational Psychology*, *108*, 114. https://doi.org/10.1037/edu0000054

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly*, *25*, 409–431. https://doi.org/10.1016/j.ecresq.2010.03.002

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*, 223–262. https://doi.org/10.1080/19345747.2012.755591

Mashburn, A., Justice, L. M., McGinty, A., & Slocum, L. (2016). The impacts of a scalable intervention on the language and literacy development of rural pre-kindergartners. *Applied Developmental Science*, *20*, 61–78. https://doi.org/10.1080/10888691.2015.1051622

Mesmer, H. A. E., & Griffith, P. L. (2005). Everybody's selling it—But just what is explicit, systematic phonics instruction? *The Reading Teacher*, *59*, 366–376. https://doi.org/10.1598/RT.59.4.6

National Center on Quality Teaching and Learning. (2015). *Language and literacy preschool curriculum consumer report*. Washington, DC: Office of Head Start, National Center on Quality Teaching and Learning. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/curriculum-report-ll.pdf

National Early Literacy Panel. (2008). *Developing early literacy*. Washington, DC: National Institute for Literacy. Retrieved from https://lincs.ed.gov/publications/pdf/NELPReport09.pdf

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly*, *46*, 249–272. https://doi.org/10.1598/RRQ.46.3.3

O'Brien, R. A., Moritz, P., Luckey, D. W., McClatchey, M. W., Ingoldsby, E. M., & Olds, D. L. (2012). Mixed methods analysis of participant attrition in the nurse-family partnership. *Prevention Science*, *13*, 219–228. https://doi.org/10.1007/s11121-012-0287-0

Organisation for Economic Development & Cooperation. (2010). *PISA 2009 at a glance*. Paris, France. Retrieved from http://www.oecd.org/pisa/46660259.pdf

Pence, K. L., Justice, L. M., & Wiggins, A. K. (2008). Preschool teachers' fidelity in implementing a comprehensive language-rich curriculum. *Language, Speech, and Hearing Services in Schools*, *39*, 329–341. https://doi.org/10.1044/0161-1461(2008/031)

Piasta, S. B., Justice, L. M., McGinty, A. S., & Kaderavek, J. N. (2012). Increasing young children's contact with print during shared reading: Longitudinal effects on literacy achievement. *Child Development*, *83*, 810–820. https://doi.org/10.1111/j.1467-8624.2012.01754.x

Piasta, S. B., Justice, L. M., McGinty, A., Mashburn, A., & Slocum, L. (2015). A comprehensive examination of preschool teachers' implementation fidelity when using a supplemental language and literacy curriculum. *Child & Youth Care Forum*, *44*, 731–755. https://doi.org/10.1007/s10566-015-9305-2

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and

children. *Journal of Educational Psychology*, *102*, 299–312. https://doi.org/10.1037/a0017763

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An ecological perspective on the transition to kindergarten: A theoretical framework to guide empirical research. *Journal of Applied Developmental Psychology*, *21*, 491–511. https://doi.org/10.1016/S0193-3973(00)00051-4

Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, *20*, 637–648.

Scott-Little, C., Kagan, S. L., & Frelow, V. S. (2006). Conceptualization of readiness and the content of early learning standards: The intersection of policy and research? *Early Childhood Research Quarterly*, *21*, 153–173. https://doi.org/10.1016/j.ecresq.2006.04.003

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education and Development*, *20*, 377–401. https://doi.org/10.1080/10409280802582795

Singal, A. G., Higgins, P. D. R., & Waljee, A. K. (2014). A primer on effectiveness and efficacy trials. *Clinical and Translational Gastroenterology*, *5*, e45. https://doi.org/10.1038/ctg.2013.13

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

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0). Retrieved from HYPERLINK "https://ies.ed.gov/ncee/wwc/FWW" http://whatworks.ed.gov.

van Steensel, R., McElvany, N., Kurvers, J., & Herppich, S. (2011). How effective are family literacy programs? Results of a meta-analysis. *Review of Educational Research*, *81*, 69–96. https://doi.org/10.3102/0034654310388819

van Tuijl, C., Leseman, P. P., & Rispens, J. (2001). Efficacy of an intensive home-based educational intervention programme for 4-to 6-year-old ethnic minority children in the Netherlands. *International Journal of Behavioral Development*, *25*, 148–159. https://doi.org/10.1080/01650250042000159

Verhallen, M. J., & Bus, A. G. (2010). Low-income immigrant pupils learning vocabulary through digital picture storybooks. *Journal of Educational Psychology*, *102*, 54–61. https://doi.org/10.1037/a0017133

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407–423. https://doi.org/10.1002/acp.1414

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, *103*, 455–469. https://doi.org/10.1037/a0023067

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology*, *91*, 261. https://doi.org/10.1037/0022-0663.91.2.261

Williams, K. T. (2007). *EVT–2: Expressive Vocabulary Test, Second Edition (EVT–2)*. San Antonio, TX: Pearson.

Appendix: Overview of RCTs With Effectiveness Characteristics

Study	Intervention	Sample size	ES phonological awareness	ES letter identification	ES vocabulary
Fischel et al. (2007)	Let's begin with the letter people; Waterford Early Reading Program	507	–	*ns*	*ns*[a]
Landry et al. (2009)	4 PD variants	1,789	.14	.15*, .47**	.15
Lonigan et al. (2011)	The literacy express preschool curriculum + PD	735	.11, .14, .17	.18, .32*, .45**[b]	.20, .27*, .32**[c]
Neuman et al. (2011)	World of words	604	–	–	*ns*
Powell et al. (2010)	PD program (coaching)	759	.22*	.32**	−.03[a]
Wasik and Hindman (2011)	Exceptional coaching for early language and literacy	541	.22*[d]	−.07[d]	.08[a,d]

Note. PD = professional development. [a]Peabody Picture Vocabulary Test. [b]Broader print knowledge. [c]Oral language Preschool Language Scale, Fourth Edition (PLS–IV). [d]The effect sizes are calculated by the authors. *$p < .05$, **$p < .01$.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Posttest Preliteracy Composite (Predicted) for Children in BAU and SPELL Conditions (Collapsed) as a Function of Children's Pretest Scores

Figure S2. Posttest Language Composite (Predicted) for Nondual Language Learners (Non-DLLs) and DLL Children in the SPELL Conditions (Collapsed) as a Function of Intervention Exposure (Mean Number of Lessons Per Child)

Table S1. Number of Child Cares, Classrooms, Educators, and Children at Two Stages in the Study (Recruitment and Posttest) for Whom Register Data Were Available

Table S2. Number of Child Cares, Classrooms, Educators, and Children at Two Stages in the Study (Pretest and Posttest) for Whom Register Data Were Available

Table S3. Attrition in Subgroups of Dual Language Learners and Categories of Parental Income and Education

Table S4. Comparison of Pretest Score for Each Subscale for All Children With Pre- and Posttests and Children With Pretest Only

Table S5. Overview of the Children Included in the External Validation Study (n = 133)

Table S6. Mean Score, Standard Deviation, and Minimum and Maximum Scores for Each Subscale and Expressive Vocabulary Test, 2nd ed. (EVT–2) and Peabody Picture Vocabulary Test, 4th ed. (PPVT–4) for 3-Year-Old Children

Table S7. Mean Score, Standard Deviation, and Minimum and Maximum Scores for Each Subscale and Expressive Vocabulary Test, 2nd ed. (EVT–2) and Peabody Picture Vocabulary Test, 4th ed. (PPVT–4) for 4-Year-Old Children

Table S8. Mean Score, Standard Deviation, and Minimum and Maximum Scores for Each Subscale and Expressive Vocabulary Test, 2nd ed. (EVT–2) and Peabody Picture Vocabulary Test, 4th ed. (PPVT–4) for 5-Year-Old Children

Table S9. Mean Score, Standard Deviation, and Minimum and Maximum Scores for Each Subscale and Expressive Vocabulary Test, 2nd ed. (EVT–2) and Peabody Picture Vocabulary Test, 4th ed. (PPVT–4) for 6-Year-Old Children

Table S10. Correlations Between Scores in Tests in Language Assessment of Children and Expressive Vocabulary Test, 2nd ed. (EVT–2) and Peabody Picture Vocabulary Test, 4th ed. (PPVT–4) for Children Aged 3–6 (n = 64–128)