Effectiveness Trial of a Language and Pre-literacy Intervention

Effective Language and Literacy Instruction:
Evaluating the Importance of Scripting and Group Size Components

Dorthe Bleses¹, Anders Højen¹, Philip S. Dale², Laura Justice³, Line Dybdal³, Shayne Piasta⁵ Justin Markussen-Brown⁶, Laila Kjærbæk⁶, & E. F. Haghish⁵,⁶,

¹TrygFondens Center for Child Research and School of Communication and Culture
²Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, USA.
³Crane Center for Early Childhood Research and Policy, The Ohio State University
⁴Rambøll Management Consulting, Aarhus, Denmark.
⁵Department of Teaching & Learning, The Ohio State University;
⁶Department of Language and Communication, University of Southern Denmark, Odense, Denmark.
⁶Clinical Epidemiology, Center of Medical Biometry and Medical Informatics, University Medical Center, Freiburg, Germany.

Correspondence about this manuscript can be sent to the corresponding author:
Dorthe Bleses,
Aarhus University
Trygfondens Centre for Child Research and School of Communication and Culture
Fuglesangs Allé 4, building 2622
8210 Aarhus V
Denmark
E-mail: bleses@au.dk
Phone: +4593505483

Acknowledgements:
This research was supported from a grant by National Board of Social Services. We wish to thanks the participating municipalities, educators, and children.
Highlights

- Language/literacy preschool intervention taken at scale had positive effect
- Effect reduced compared to typical effects in efficacy trials
- Larger effect when activities were not aligned the sequence and scope but decided by educators
- Effect not influenced by child group size or children being/not being at risk
- Significant association between exposure and effect
Abstract

Identification of intervention program components most strongly associated with children’s outcomes is essential for designing programs which can be taken to scale. In this effectiveness study, a population-representative sample of 5,436 3-6 year old Danish children from 154 daycare centers participated in a cluster-randomized evaluation of three variations of a language-literacy focused curriculum (LEAP) comprised of 40 twice-weekly 30-minute lessons. LEAP-LARGE and LEAP-SMALL conditions were based on a scope and sequence of objectives plus scripted lessons, provided to whole-class and small groups, respectively. In LEAP-OPEN, the scope and sequence was provided, but educators determined the instructional activities for each of 40 lessons. A business-as-usual (BAU) condition served as the control. Overall, the largest effect sizes for children’s language and emergent literacy outcomes were found for LEAP-OPEN, although the other two LEAP conditions had positive effects for literacy outcomes. Analysis of moderation effects showed no effects for children’s socioeconomic status, but interactive effects for immigrant children: specifically, non-Danish children benefitted for language outcomes. Finally, there was a significant association between children’s amount of exposure to the program and outcomes for both measures. This study provides evidence that systematic, explicit language and literacy interventions implemented at-scale do achieve impacts of practical significance even though implementation and effect size are somewhat reduced.

Keywords (max 6): Language and literacy intervention; randomized controlled trial; large-scale effectiveness study, intervention components (group size and sequence and scope with aligned activities)
Effective Language and Pre-literacy Instruction:

Evaluating the Importance of Scripting and Group Size Components

The language and literacy skills of preschool-aged children are strongly predictive of their future skills as readers and writers (see National Early Literacy Panel [NELP], 2008). Consequently, researchers have devoted considerable attention to designing and testing systematic, explicit interventions for preschoolers to improve their language and literacy skills in such areas as phonological awareness, print knowledge, and vocabulary knowledge (e.g., Justice, McGinty, Piasta, Kaderavek, & Fan, 2010; Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013; Lundberg, Frost, & Petersen, 1988). Systematic, explicit language and literacy interventions typically provide a scope and sequence of objectives and scripted lessons with aligned activities for educators to follow to directly teach to these objectives, often as a whole-class activity (e.g. Bierman et al., 2008; DeBaryshe & Gorecki, 2007; Gonzalez et al., 2011; Justice, McGinty, Piasta, Kaderavek, & Fan, X., 2010). Despite the encouraging body of evidence for the effects of systematic, explicit language and literacy intervention on young children’s development of these skills, our understanding of whether and to what extent these interventions work when taken to scale is limited. For instance, it is unclear whether practitioners can adopt the critical components of language and literacy intervention with fidelity, such as adhering to scripted lessons and implementing instruction in a large-group setting. There is a complementary need to identify the individual components of systematic, explicit language and literacy curricula that are most strongly associated with improved child outcomes. The focus of the present study was to determine the effects associated with two components commonly seen in systematic explicit language and literacy interventions (i.e., group size and scripted lessons with aligned activities) on child outcomes when implemented at-scale. The main aim is to obtain knowledge about which elements are most effective children’s outcome to be able to focus on achieving high implementation fidelity to
these elements in order to increase the overall effectiveness of language and pre-literacy interventions. Additionally, we wanted to explore if individual child-level difference (i.e., background characteristics and individual intervention exposure) moderated children’s outcomes.

**Efficacy and effectiveness trials of language and literacy interventions**

The majority of rigorous, published studies evaluating the impact of interventions have been conducted under relatively optimal conditions, including high levels of implementer training, careful monitoring of compliance, and stringent selection of participants (Singal, Higgins, & Waljee, 2014). These studies have been labelled *efficacy* trials. Based on the evidence accumulated largely from efficacy trials, implementation of systematic, explicit language and literacy interventions leads to improvements in targeted skills of the magnitude of 0.25 to 0.87 standard deviation units (NELP, 2008). In contrast, *effectiveness* studies evaluate the impact of interventions carried out under relatively routine conditions.

A related concept, *scale-up*, refers to the deliberate expansion of interventions so that they might be accessed by a larger, more diverse population (O’Donnell, 2008). Interventions that are appropriate for scale-up are those that exhibit two core characteristics: (1) they yield positive effects on the outcome of interest as typically evidenced in efficacy studies, and (2) they can be implemented with fidelity (O’Donnell, 2008). Systematic, explicit language and literacy interventions therefore generally meet the first characteristic of interventions that are appropriate for scale-up. The well-documented positive impacts of systematic, explicit language and literacy interventions for young children suggests that there may be value in taking such interventions to scale. In many cases, however, effects observed when interventions are taken to scale are attenuated from those seen in efficacy studies.

A notable limitation in the current research literature is that very few large-scale investigations have sought to determine the effectiveness of implementing systematic, explicit language
and literacy interventions carried out under routine conditions with large groups of early childhood educators (Landry, Anthony, Swank, & Monseque-Bailey, 2009, 2011). Such studies would address the important issues of both effectiveness and scale-up. A notable exception is the work of Landry and colleagues, who conducted several effectiveness studies involving educators working in settings that served primarily at-risk children, including children reared in poverty.

Author and colleagues, on the other hand, conducted a scale-up study involving a heterogeneous sample of children e.g., Authors, 2014; under review. This study evaluated the effectiveness of an experimental systematic, explicit language and literacy intervention that was implemented across Denmark under real-life conditions. The intervention, titled SPELL (Structured Preschool Efforts in Language and Literacy), was implemented for 6,483 children enrolled in Danish daycares that served a general, unselected population of Danish 3- to 6-year-olds. In comparison to business-as-usual daycare, children exposed to SPELL exhibited significant gains in several literacy skills, with the size of the effect depending on the domain (Autours, 2014; Authors, under review). Interestingly, the effects obtained in this cluster-randomized, scale-up study are comparable with those effectiveness trials that have focused on at-risk populations within the United States (e.g. Landry et al., 2009; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011; Wasik & Hindman, 2011), suggesting that scaling-up systematic, explicit language and literacy interventions may yield benefits to all children, not only those from high-risk backgrounds.

To this end, and given the heterogeneity of their sample, Author and colleagues (2014; under review) also investigated the differential effects of the systematic, explicit language and literacy intervention for subgroups of children, in order to determine whether certain groups particularly benefitted from the intervention. In the Author study, neither socioeconomic status (SES, indexed as maternal education) nor dual language learner (DLL) status significantly moderated the intervention effects, suggesting that intervention effects were not constrained by these child-level attributes, a conclusion that differs from that of recent reviews based on efficacy studies (Buysse,
Such studies suggest that children from low-SES backgrounds and/or who are DLLs gain less from language and literacy interventions than children from more advantaged backgrounds. Results like these suggest, as would be anticipated, that taking systematic explicit language and pre-literacy 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. The reduced effect sizes may be related to the latter characteristics of interventions mentioned above, that is, that they cannot be implemented with sufficient fidelity and this characteristic has been more elusive (see Pence, Justice, & Wiggins, 2008).

**Fidelity of implementation**

Within the early language and literacy intervention literature, several studies have found that educators’ fidelity of implementation for important aspects of the intervention, such as delivery of the desired number of lessons and adherence to the particular components of the intervention (see Darrow, 2013) can be quite low, even when educators receive considerable support from research staff or coaches. For instance, an evaluation of educators’ fidelity of implementation to the language and literacy components of the MyTeachingPartner intervention, as implemented by 154 preschool educators, found that only 45% of educators achieved the desired dosage per week, with wide variability (Hamre et al., 2010). This is despite the fact that educators in this study received ongoing implementation supports (e.g., incentives, materials) as well as coaching. An earlier study by Pence and colleagues similarly found that preschool educators who were implementing a multi-component language intervention used very few of the language-facilitating techniques that were deemed a central component of the intervention (Pence et al., 2008). In this study, the educators were closely monitored by research staff and received ongoing training and coaching in use of these techniques.
It is important to note that most studies of preschool educators’ implementation of language and literacy interventions have involved educators who are actively involved in a research study; we know even less about what happens when educators are separate, physically and organizationally, from research efforts and receive limited support from research staff. A notable exception is a recent study by Bierman and colleagues (2013), in which 37 preschool educators participated in an efficacy study in which they implemented a complex set of language, literacy, and social-emotional interventions in their classrooms for an academic year; educators received all intervention materials and weekly coaching to promote implementation. These supports were withdrawn after the initial year of investigation, but educators were encouraged to continue implementing the interventions. Study results showed that educators’ implementation fidelity for the language and literacy intervention declined significantly from the first to the second year, suggesting that intervention fidelity may be especially low when educators are not actively coached in intervention implementation.

This point is supported in the work of Author and colleagues, who conducted the Danish scaling-up trial referenced previously (Authors, 2014, under review). In that trial, researchers maintained a distance from the involved educators and fidelity was not reinforced, only monitored. On average, educators implemented 75% of intervention lessons, but only 10% of educators completed all 40 intervention lessons. Interestingly, educators’ level of implementation fidelity accounted for 9% to 19% of the variance in children’s literacy and language gains, respectively. This finding is similar to that of Wasik and Hindman (2011), who found substantial variation in fidelity across educators when implementing a language intervention, which explained 28% of the variation in child outcome for phonological sensitivity and 29% for vocabulary. The results suggest that intervention impact is reduced as a function of lower implementation fidelity.

Many factors might account for these low levels of fidelity by early childhood educators. Given the relations between implementation fidelity and child outcomes, this is an important
topic for research. Some have proposed that intervention fidelity is compromised when the instructional practices educators are being asked to adopt do not readily fit within the educators’ belief systems or their standard practice of education and care (McDonald et al., 2006). For instance, educators may be reluctant to use whole-group instruction in early education settings, as it may not be perceived as a developmentally appropriate practice (File & Gullo, 2002); however, whole-group instruction is an instructional practice used in many language and literacy interventions (Assel, Landry, Swank, & Gunnewig, 2007; Justice, Logan, Kaderavek, & Dynia, 2015; McGinty, Breit-Smith, Fan, Justice, & Kaderavek, 2011; VanDerHeyden, Snyder, Brussard, & Ramsdell, 2008; Wasik, Bond, & Hindman, 2006).

The extent to which specific components of language and literacy intervention may compromise fidelity and/or attenuate effects of the intervention on children’s outcomes is currently unclear, yet is an important and necessary focus of recent investigations (Lonigan, Purpura, Wilson, & Clancy-Menchetti, 2013).

**Evaluating the components of interventions**

Many preschool-focused language and literacy interventions are multi-component, including specification of a scope and sequence delineating a set of specific targeted skills organized over the period of instruction, and instructional practices (i.e., techniques and aligned activities). In many interventions, the scope and sequence and the aligned instructional practices are manualized via *scripted lessons* that educators implement with their students. These scripts serve to identify the specific instructional techniques and activities within each lesson, with the techniques and activities serving to deliver the scope and sequence of objectives to students (e.g., Assel et al., 2007; Bierman et al., 2008; Hamre et al., 2010; Justice et al., 2010). In addition to being a core component of the intervention itself, scripted lessons are often an important aspect of the research design, with teachers’ adherence providing an index to teachers’ fidelity of implementation.
Nonetheless, the use of scripted lessons may potentially attenuate teachers’ implementation fidelity as well as child outcomes, as educators may feel that using scripted lessons precludes being responsive to children’s needs or that their autonomy as professionals is undermined (Achinstein & Ogawa, 2006). Despite the prominence of scripted lessons in language and literacy interventions in early education settings, the potentially causal role they play as an intervention component has not, to our knowledge, been investigated.

Another frequent component of language and literacy intervention is the delivery of the intervention, or at least parts of it, as a whole-class activity. It is possible, however, that this intervention component may attenuate intervention effects on children relative to use of individual or small-group instruction. As we noted previously, the use of teacher-directed, large-group instruction in early childhood settings has long been seen as developmentally inappropriate (Burts et al., 1990), and may compromise young children’s learning. In a conceptual review, Wasik (2008) maintain that there are significant benefits to small group instruction as they afford the opportunity for teachers to devote more attention to children and therefore more opportunities to learning and developing for young children. Small group instruction are therefore better than whole groups to teach content if certain guidelines are followed, i.e., the group size should not exceed five children, small groups should be intentionally organized the content presented in small groups, teachers need to play an active role. However, Marulis and Neuman (2010) conducted a meta-analysis of vocabulary intervention studies for young children, which included differentiating effects as a function of individual, small group, and large group interventions. This study did not find that group size was associated with child outcomes in vocabulary. Nonetheless, recent studies do not necessarily support this point at least in children at risk (see e.g. Kruse, Spencer, Olszewski, & Goldstein, 2015; Lonigan & Phillips, 2016). However, in the broader context of language and literacy interventions, it is unclear whether the use of whole-group programming as a frequently observed component is as effective as the use of smaller groupings.
The present study

The present study was designed to test the effects of two specific components of a systematic, explicit language and literacy intervention, namely, scripted lessons that serve to manualize the scope and sequence of intervention and group size. The importance of these components to language and literacy intervention was tested in an at-scale effectiveness trial in Denmark, with the following characteristics: (1) participation by eight representative municipalities from across the country; (2) obligatory participation by randomly selected daycares, as the intervention was implemented as a municipality-sanctioned program; (3) enrollment of all children in the selected daycares into the trial; (4) implementation of the intervention by practicing educators; and (5) little to no active enforcement of implementation among participating educators.

The trial described herein was conducted in parallel but independently of another scale-up trial (Authors, 2014, under review) designed to test the previously mentioned SPELL intervention within the Danish daycare context. The Danish daycare system is distinguished from the American system due to nearly universal enrollment of children between 1 and 5 years. Extensive investment in daycare by the local and federal government is consistent with the design of the Danish welfare state, in which nearly all women and men are employed full-time. Thus provision of child-care is crucial and municipalities are obliged to ensure availability of daycare (Datta Gupta & Simonsen, 2013). For daycares serving children three to five years of age, the educator-child ratio is relatively low at one educator to about seven children whereas in the US and Canada the corresponding ratio is 12:1 (Dalsgaard, Nøhr, & Jordan, 2014. Approximately 60% of educators have a 3.5 year pedagogical bachelor degree (Dalsgaard, Nøhr, & Jordan, 2014). Danish daycares typically use child-oriented pedagogies and educators are trained to take a holistic approach to pedagogy (Bauchmüller, Görtz & Rasmussen, 2014). Encouraging democratic values and social skills are cornerstones of practice, whereas early academic goal setting is generally treated warily. There are no specific learning goals mandated by the government except for general learning
planns, and a rigorous, empirical basis for practice has not been available, leading to the conduct of two large-scale trials (the present study and the SPELL trial) to build this base.

Specifically, the present trial featured an empirical evaluation of a play activity-based approach to providing explicit language and literacy intervention, which was designed to emulate the play-based instructional activities typically used in standard practice of early education (such as storytelling using pictures, nursery thymes, memory games and motor activities) in Denmark. LEAP (Language Education Activities for Preschoolers) uses play-based activities to systematically and explicitly target 23 language and literacy learning objectives, identical to those of SPELL and derived from a curriculum supplement developed and tested in the United States (see Justice et al., 2010; Justice and McGinty, 2012) in a whole-class routine. In conducting this trial, we not only examined the effects of LEAP on children’s language and literacy skills, relative to business-as-usual (BAU) daycare practice, but we also investigated the effects of two core components of language and literacy interventions on child outcomes, namely the grouping practices employed (large, small) and the use of scripting with aligned activities (scripted aligned, unscripted unaligned). With four conditions in total, we created counterfactuals to assess the effects of grouping and scripting with aligned activities. First, in a LEAP-LARGE condition, educators implemented the LEAP intervention in a whole-group format in their classrooms (similar to most US interventions), with a LEAP-SMALL condition providing a counterfactual in which the LEAP intervention was implemented in small groups. We speculated that child outcomes would be heightened in the LEAP-SMALL condition versus the LEAP-LARGE condition, given the plausible assumption that both the quantity and quality of educator-child interactions would be increased in small groups. Second, in a LEAP-OPEN condition, educators implemented small group instruction where they were asked to explicitly target the same 23 language and literacy learning objectives as in LEAP-LARGE and LEAP-SMALL (i.e. they should follow a scope and sequence of instruction) the these educators were not provided with scripted lessons with aligned
activities. Instead they were instructed to address the learning objectives within twice-weekly small-group activities of their own choice. This provides a counterfactual to the LEAP-SMALL condition and allows us to examine the effects of scripting with aligned activities on child outcomes. We speculated that child outcomes would be enhanced in the LEAP-OPEN condition, relative to the LEAP-SMALL condition, as educators would be able to better integrate the intervention with other educational activities and use children’s current engagement and interest in topics. Moreover, this may support educators’ sense of autonomy as a professional. There are, however, plausible reasons the results might go in the other direction. In examining child outcomes relative to these four conditions, we also examined whether impacts were moderated by specific characteristics of children and of intervention exposure. In total, three questions were addressed: (1) To what extent does educators’ implementation of three variations of LEAP, one featuring large-group instruction (LEAP-LARGE), one featuring small-group instruction (LEAP-SMALL) and the last featuring no scripted lessons and no aligned activities within a small-group context (LEAP-OPEN), influence the language and literacy skills of children relative to business-as-usual (BAU) instruction? (2) To what extent are the impacts of LEAP conditional on children’s background characteristics (DLL, household SES)? and (3) To what extent is children’s intervention exposure to SPELL associated with language and pre-literacy outcomes for children as a function of different child characteristics?

**Method**

**Participants**

**Daycare program selection and assignment.** For the present study, all Danish municipalities were invited for participation. Denmark has a local government system with 98 municipalities, which handle most welfare tasks including the administration of daycare facilities. Daycare is offered universally across the country, and the local council of the authority in each municipality must provide guaranteed daycare availability to all children aged 26 weeks to school age.
An estimated 97% of children between 3 and 5 years of age are enrolled in daycare (Danish Ministry for Social Affairs, 2015). The centers are generally situated in residential areas across Denmark, thus their socio-economic composition is solely determined by the composition of the neighborhood. Daycares vary greatly in size (range 11-80 children) but most daycares enroll an average of 40 to 50 children (Nøhr, Dalsgaard, Kloppenborg, Meldgaard, & Bækggaard, 2012). Most daycares are organized into classrooms (approximately 15-20 children on average) but some daycares use activity centers as the main principle of organization.

For this study, eight municipalities were purposefully selected from among the 20 municipalities that actively applied for participation in the study; selection of the municipalities focused on ensuring that participating daycares were diverse with respect to geography and children’s background characteristics. Across two consecutive waves, a total of 156 daycare centers were recruited from these eight Danish municipalities. Following recruitment, all daycare centers in each municipality were stratified on the basis of information from Statistics Denmark on ethnicity, family structure (single parent or not), parental education, parental income, and use of social services; after stratification, daycares were randomly assigned to four conditions. All classrooms and educators within the selected daycares were included in the study and assigned to the condition of their center. Two daycares withdrew just prior to intervention. The remaining 154 daycares were slightly unevenly distributed across the four conditions as shown in Table 1. Data for all randomized daycares, educators, and children for whom background data could be obtained (n = 7,209) are shown in Table 1. All classrooms included children between the ages of 3 and 6 years.

To determine whether children in the four conditions differed at the start of the study, background variables and pretest scores across the four conditions were examined using hierarchical linear models (see Analytic Strategy). The results of these models revealed that daycares, educators and children across the four conditions did not differ significantly on any of the background
variables or pretest scores (all \( ps > .096 \), uncorrected). Further, hi-square tests indicates that there were no significant differences across the different SPELL conditions in terms of children’s gender, age, DLL status, mothers educational background (low, high) and income (low, high, all p values > .05).

This study had some missing data as is expected in large-scale trials. The primary child-level measures of interest were assessments administered by the educators in each classroom, as seen in similar large-scale studies (Mashburn et al., 2012). Pretests were obtained for 6,675 children. First, some data were missing due to attrition of childcares. Between pretest and posttest, nine daycares withdrew. Eight withdrew due to leadership or staff transitions and difficulties. In a ninth daycare, posttest data were never received by the project management. Because child-level assessments were carried out by the educators, it was not possible to obtain comparable posttest assessments after this attrition, as would be ideal in a true intent-to-treat design. Second, pretests and posttests were missing for some children, due to illnesses or other absences at each time-point resulting in a final sample of 5,436 children with pretest, posttest, and age information in the four conditions (BAU, \( n = 1,436 \); LEAP-LARGE, \( n = 1,229 \), LEAP-SMALL, \( n = 1,378 \); LEAP-OPEN, \( n = 1,393 \)). Total attrition percentages of children between pre- and posttests for each condition ranged from 16% to 21% (BAU, 16%; LEAP-OPEN, 18%, LEAP-SMALL, 20%; LEAP-LARGE, 21%). The number of children tested on each subscale varied depending on 1) age-related administration of subscales, and 2) failure of a few children to complete all subscales. All children who had both pre- and posttest scores were included in analyses. As standard software for multiple imputation does not allow a five-level nested design, which characterized the nested data structure for this study, we included in each of our analyses only children for whom information on all variables involved in the analysis were available.

**Procedures**
The primary procedures concerned implementation of the three LEAP conditions, or maintenance of typical practice in the BAU, as well as measurement of educators’ implementation fidelity, children’s treatment exposure, and children’s language and literacy skills at pretest and posttest.

**LEAP conditions (LEAP-OPEN, LEAP-LARGE, LEAP-SMALL).** Educators assigned to implement one of the three planned variations of the LEAP intervention did so as a requirement of their municipality administration and thus were not self-selected as intervention implementers. Across all three conditions, the LEAP educators were asked to implement an experimental intervention for a 20-week period, follow a scope and sequence of language- and literacy-focused instruction, and complete a variety of implementation fidelity tools (see Justice et al., 2010; Justice & McGinty, 2012). Given that two of the LEAP conditions, namely LEAP-SMALL and LEAP-OPEN, involved delivering the intervention in small groups, the educators within each of these classrooms organized the children into small groups of up to six children based on their knowledge of children, including knowledge regarding age, prior skills, and relationships with other children. Given that a typical classroom contained about 16 children and two educators, on average there were four small groups within each classroom. The two educators in each classroom were randomly assigned to two groups of children within a classroom, and provided the LEAP intervention to these intact groups for the duration of the 20-week period. In LEAP-LARGE, all children in the classroom participated in the same group, and each lesson was implemented by two educators in a classroom.

LEAP-SMALL and LEAP-LARGE involved educators’ implementation of 40 scripted 30-minute lessons twice per week to either children in small groups (LEAP-SMALL; mean group size = 4.9, SD = 1.4) or to the entire class (LEAP-LARGE; mean group size = 16.0, SD = 4.7). The lessons addressed four learning domains (phonological awareness, vocabulary knowledge,
print knowledge and narrative ability) and a total of 23 objectives ordered over the 20-week period. The specific domains and objectives are presented in the Appendix in both English and Danish. To guide implementation of this scope and sequence, educators were provided soft-scripted lesson plans to follow during each lesson within the context of a small-group or large-group activity. The lesson plans featured instructional play-based language and literacy activities that were carefully developed for the purposes of this study by the study team. The activities included memory games, nursery rhymes, children’s songs, storytelling using pictures, and various activities that involved moving around in the daycare. Each type of activity was repeated four to five times across the 40 lessons to support the four learning domains. All types of activities were piloted in 16 daycares and educators provided oral and written feedback to help refine activities prior to implementation.

The lesson plans were crafted so that each lesson emphasizes one objective for each of two learning domains (e.g., phonological awareness and print knowledge). Educators were provided a set of materials required to implement these lessons, to include a suitcase, picture cards, and a “booklet” for each child with some printed material, but most lessons utilized toys or other material prevalent in all daycares. Each lesson plan included two components: (1) the LEAP lesson plan itself, which soft-scripted a sequence of step-by-step instructions for the aligned language and literacy activity as well as suggested language that educators could use to support children’s learning during each activity, and (2) a ‘Learners’ Ladder,’ which presented specific scaffolding strategies that educators could use for children who found a given lesson too easy or too difficult (for more details, see Justice & McGinty, 2012).

LEAP-OPEN differed from the other two variations in that the educators were not provided soft-scripted lessons to use within the 40 lessons implemented over the 20-week period of intervention. Rather, these educators were provided the scope and sequence of instruction used in
Effective Language and Pre-literacy Instruction 18

LEAP-OPEN and LEAP-LARGE and were instructed to address all of them, yet they had the autonomy to decide how many of the learning domains and objectives they would address in each lesson and in which activity. As in LEAP-SMALL, the lessons were implemented by the educators in a small-group setting at an intensity of two per week lasting approximately 30 minute per lesson. To facilitate implementation of this condition, the educators received a short manual in which the learning domains, learning objectives, and basic principles of the Learner’s Ladder were described.

To support implementation of each lesson, educators in all three conditions completed implementation notes after each lesson on a web-based platform (see Measures). Also, educators tracked individual children’s progress towards the 23 LEAP objectives on the web-based platform three times during the 20-week-study period (after lessons 6, 20, and 36) using an informal assessment on which educators indicated whether the child never, sometimes, or often demonstrated the skill.

BAU represented the control condition in this study. Educators in the BAU condition were instructed to carry out their usual practice, which was generally very distinct from the explicit language and literacy instructional approach utilized in the LEAP variations as learning in social interactions rather than structured instructional situations are preferred in Danish daycares (Broström, Johansson, Sandberg, & Frøkjær, 2012).

Implementation supports. Educators received all manuals and materials needed to implement the intervention. Educators in all three LEAP conditions received a two-day training workshop (14 hours) which incorporated elements of practice-focused professional development in the coursework by using many instructional videos. Educators were trained in identifying the four learning domains and the 23 learning objectives within each of the domains and how to reference these explicitly during activities. The educators were also trained in identifying and using the Learner’s Ladder (Justice & McGinty, 2012). However, only educators in LEAP-SMALL and
LEAP-LARGE were introduced to the program sequence and specific activities. During those times, educators in LEAP-OPEN were grouped to discuss and plan how they wanted to design and plan the intervention. In the BAU comparison condition, educators continued to provide the types of instruction typical in Danish daycares. To control for Hawthorne effects, educators in that condition received a one-day training workshop on social inclusion, which was relevant to their daily educational practice.

**Implementation fidelity.** Two types of information were used to track treatment fidelity in all the experimental conditions: (1) self-reporting based on completion of implementation notes after each lesson on the web-based platform and (2) observational ratings. The implementation notes also documented child exposure, defined as number of completed lessons a child was exposed to (sometimes referred to as program differentiation, cf. Darrow, 2013), and adherence to the intervention, defined as the extent to which the intervention components (use of intervention material and activities, addressing of learning objectives, use of scaffolding strategies) were implemented. Finally, educators indicated the performance of individual children in their group and their engagement in lesson activities. In LEAP-OPEN, for each lesson the educators also had to report which learning domain and objectives they addressed, the type of activities they used, and the extent to which they worked explicitly with the prescribed learning domains and objectives during the activities.

Observational ratings were used to assess adherence to the critical elements of the intervention and participant responsiveness based on video recordings of lessons submitted by the educators in the beginning, middle, and end of the 20-week program. Research staff coded the videos submitted by the educators on the basis of a checklist developed for this specific purpose. Approximately 10% of videos were coded by two master-coders to assess interrater agreement. Agreement between a master-coder and the research staff ranged from 77% to 96% (only one comparison below 80%).
Table 2 provides descriptive data regarding implementation fidelity, based on educator’s submission of implementation notes as well as the observations. Across the three conditions, educators implemented between 73% and 88% of lessons. Comparing the number of completed lessons across conditions, there was a significant difference between the conditions, $\chi^2 (2, n=645) = 8.46, p = .014$ (more lessons were carried out in the LEAP-LARGE condition). Regarding children’s exposure to the conditions, recorded for each child by their educator, children were exposed to between 25 and 29 of the 40 lessons, on average. There was no significant difference across conditions in terms of exposure for individual children, $\chi^2 (2, n=3,982) = 4.55, p = .10$.

The adherence to the core elements of the scripted lessons (use of materials, activities and sequence and scope) was relatively high based on the self-reported measures, with educators reporting to use about 78% to 79% of lesson elements. Observational data analyzed by research staff showed adherence to be slightly lower.

Finally, individual children’s general performance and engagement in lessons was rated significantly higher by educators in LEAP-OPEN compared to educators in LEAP-SMALL ($\chi^2 [1, N=3,407] = 40.6, p = <.001$ and $\chi^2 [1, N=3,407] = 31.4, p = <.001$) (individual children’s engagement were not reported for LEAP-LARGE).

Measures

The language and emergent literacy skills of the children were assessed pre- and post-intervention using the Language Assessment of Children: 3-6 (LA; Bleses, Vach, Jørgensen, & Worm, 2010) instrument. LA is administered by educators in daycares in almost all Danish municipalities as part of a national screening program, including all municipalities participating in the present study. Two age-dependent versions of the instrument were administered by educators in the participating daycares. Four subscales were administered to 3-year-old children: Sound Discrimination (of initial phonemes in words, max score 16), Vocabulary (expressive, max score 76), Comprehension (of words and complex sentences, max score 12), and Communication
Six subscales were administered to 4- to 6-year-old children: Rhyme Detection (max score 17), Deletion (max score 20), Letter Identification (max score 12), Vocabulary (expressive, max score 40), Comprehension (of words and complex sentences, max score 27), and Communication (questionnaire with communicative strategies, max score 76). With the exception of Communication, children were directly assessed via picture identification or picture elicitation tasks.

For this study, two composite measures were derived from the seven subscales: a Pre-literacy Composite based on averaging standardized scores for Sound Discrimination, Rhyme Detection, Deletion, and Letter Identification (Pre-literacy outcome for the three year old children was Sound Discrimination only) and a Language Composite based on averaging standardized scores for Vocabulary, Comprehension, and Communication. Standardized scores were based on age and gender specific norms derived from the pretest measurements. Internal consistency coefficients (Cronbach’s alpha) for subscales across subscales were ranged .75 to 91; correlations between subscales ranged .25 to .70. The concurrent correlations of the language subscales with The Peabody Picture Vocabulary Test-4 (PPVT-4, Dunn and Dunn, 2012) were .55 for Vocabulary and .57 for Comprehension; correlations with the Expressive Vocabulary Test 2 (EVT-2, Williams, 2007) were .42 for Vocabulary and .39 for Comprehension. For the pre-literacy subscales, the correlations with PPVT4 were .32 for Rhyme Detection, .42 for Deletion, and .39 for Letter Identification; the correlations with EVT2 were .18 for rhyme, .29 for Deletion, and .33 Letter Identification (Bleses, Højen, Dale, & Makransky, 2016).

In interpreting the effects of the intervention in the following sections, it is worth mentioning that some of the subscales exhibited ceiling effects 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 subscale 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 Identification (score at 75th percentile = 11, maximum = 12), Rhyme (score at 90th percentile = 16, maximum = 17), and a less pronounced ceiling effect for Vocabulary (score at 90th percentile = 69, maximum = 76). Given these ceiling effects on the pretests, a proportion of the children (10% to 75% depending on subscale) had no possibility of showing improvement on posttests, which may limit estimated effect sizes.

**Analytic Strategy**

The data collected in the LEAP trial were nested in a complex way, with children nested in either small groups (LEAP-SMALL and LEAP-OPEN) or large groups (LEAP-LARGE), educators (who delivered the intervention), classrooms, and center. Therefore the research questions were addressed using 5-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. The Level 4 (classroom) and 5 (center) variables were treated as random effects for all three conditions; for the Level 2 (group) and 3 (educator) variables, however, they were treated as random effects for the two intervention conditions only, as children in the LEAP-LARGE and the BAU condition were not assigned to groups or to educators. Intraclass correlations (ICCs) provide evidence for variance at each of the five levels. At the center level, ICCs ranged from 0 on Language Composite to .34 on Sound Discrimination; at the classroom level, ICCs ranged from 0 on Sound Discrimination to .12 on Communication; at the educator level, ICCs ranged from .05 on Deletion to .23 on Sound Discrimination; at the group level, ICCs ranged from 0 on Letter identification to .09 on Communication; finally, at the child level, ICCs ranged from .43 on Sound Discrimination to .89 on Letter Identification. These values indicate that the majority of variance lay between children but that each level contributed some variation.

Specific covariates were included in the models for each of the research question. For all relevant research questions, the pretest value corresponding to the posttest value of interest was
always included as a covariate. When the Comprehension scale was included, the version of LA was included as a covariate to take into account the different number of items in the two age-dependent versions. Due to lack of information about maternal education for many DLL children in Statistics Denmark (data on maternal education is only registered for those mothers who complete their schooling in Denmark), DLL status was considered as an additional category when maternal education was entered into a model as a covariate, such that the effect of maternal education was estimated only for the Danish children.

Cohen’s $d$ effect sizes were calculated by dividing the HLM effect estimates by the standard deviation of the individual change scores (Cohen, 1988). The six pairwise comparison $p$-values were based on the Kramer-Tukey method 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.

**Results**

**Main Effects**

To address the first research question (namely, to what extent does educators’ implementation of three variations of LEAP influenced the language and literacy skills of children relative to BAU instruction), we estimated nine separate hierarchical linear models (HLMs), one for each outcome variable and the two composites. The following fixed effect covariates were included at each level: at level 1 (child-level), the corresponding pretest score, age, subtest version (depending on age; only relevant for comprehension subtest), and gender, and at level 5 (daycare-level), the condition to which the daycare was assigned, the cohort in which the daycare participated, and the municipality of the daycare. Results for the nine HLM analyses appear in Table 3 as well as Cohen’s $d$ effect size estimates for each of the three intervention conditions compared to BAU.

The results presented in Table 3 show that the LEAP-OPEN condition, in which educators composed their own lessons to address the given learning objectives, resulted in significant effect
sizes for all outcome measures except Communication as compared to BAU. The other two conditions (LEAP-LARGE and LEAP-SMALL), in which educators were provided with soft-scripted lessons with aligned activities in either large-group or small-group settings, also showed differences relative to the BAU. For the LEAP-LARGE and BAU contrast, results showed that children in the LEAP-LARGE condition outperformed those receiving BAU only on the Rhyme Detection ($d = 0.18$) and Letter Identification ($d = 0.25$) subtests and on the Pre-literacy Composite ($d = 0.20$). For the LEAP-OPEN and BAU contrast, results showed that children in the LEAP-OPEN condition outperformed those receiving BAU on all subscales except for Sound Discrimination and Communication (Rhyme Detection, $d = 0.38$; Deletion, $d = 0.30$; Letter Identification, $d = 0.31$; Vocabulary, $d = 0.33$; Comprehension, $d = 0.30$) and the two composites (Pre-literacy Composite, $d = 0.32$; Language Composite, $d = 0.30$).

In addition to differences between the BAU condition and the three intervention conditions, there were several significant differences between the intervention conditions. LEAP-OPEN yielded significantly larger effects than LEAP-SMALL on Vocabulary ($p < .001$), Comprehension ($p = 0.004$), Communication ($p = .006$), the Pre-literacy Composite ($p = .024$), and the Language Composite ($p < .001$). LEAP-OPEN yielded significantly larger effects than LEAP-LARGE on Rhyme ($p = .014$), Vocabulary ($p < .001$), Comprehension ($p < 0.001$), and the Language Composite ($p < .001$). No comparisons between LEAP-SMALL and LEAP-LARGE were significantly different.

**Effect Moderation by Child Characteristics**

The second research question examined the extent to which certain child-level characteristics (maternal education and DLL status). To address issues of moderation, we focused our analyses only on the children in the LEAP-OPEN condition relative to the BAU, as this experimental condition consistently yielded the largest effects relative to all three counterfactuals. In addition,
we examined moderation only as it related to the two composite outcomes (Pre-literacy Composite, Language Composite) for the purpose of parsimony. In these models, the child-level covariates of interest (maternal education, DLL status) were included both as a main effect and as interactions with condition. The continuous variables were grand mean centered. Additional covariates included in the models included children’s pretest scores, age, and gender.

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 LEAP is the difference between the BAU condition and the experimental condition 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 these variables change if the child was assigned to BAU versus the intervention condition. For the interaction including the DLL variable, the interaction can be best interpreted as the change of the LEAP effect for a child with non-immigrant versus to a child with and immigrant background.

Among the variables included in the first model, for which the Pre-literacy Composite served as the outcome variable, only children’s pretest scores (p < .001) and age (p < .001) served as unique, significant predictors of the outcome; in addition, condition served as a unique, significant predictor of the outcome (p < .001). In terms of the interaction between the covariates and condition, only the treatment by age interaction was significant (p < .001), as shown in Figure 1. Specifically, the results showed that older children gained more from the LEAP-OPEN intervention than younger children did.

Of the variables included in the second model, for which the Language Composite served as the outcome variable, pretest (p < 0.001), age (p < .001), DLL status (p < .001) and maternal education (p = .001) were independent, significant predictors. Neither condition nor any of the
other variables significantly interacted with children’s outcomes. Taken together, these interaction analyses showed that none of the examined child-level variables interacted with the condition, with the exception of age for the pre-literacy outcome.

**Impact of Intervention Exposure for Child Outcomes**

Of final concern was examining the extent to which child exposure (mean number of lessons in which the child participated) predicted outcomes as a function of different child characteristics; here, two additional HLMs were estimated, one for each of the posttest composites (Pre-literacy Composite, Language Composite). In these models, predictors included were the corresponding pretest, age, gender, DLL status, and maternal education as well as the interaction terms for intervention exposure X DLL status and intervention exposure X maternal education. The effect of intervention exposure represents the increase in the posttest score associated with 10 more intervention lessons (for a non-DLL boy with average age, average pretest score, and average mother’s education). The interaction terms describe how this number changes considering a non-DLL versus DLL child (intervention exposure X DLL status) or if we increase maternal education by one level (intervention exposure by maternal education).

Table 5 provides the unstandardized coefficients (B), standard errors, and *p*-values for each of the predictor variables and interaction terms. In the first model in which the Pre-literacy Composite served as the outcome variable, all variables but gender (boy) served as an independent and significant predictors of the outcome variable. Neither Intervention Exposure X DLL status (*p* = .514) nor Intervention Exposure X Maternal Education (*p* = .101) interacted significantly to predict the outcome variable.

In the second model, in which the Language Composite outcome composite served as the outcome variable, children’s pretest scores (*p* < .001), age (*p* < .001), DLL status (*p* < .001), and intervention exposure (*p* < .001) served as independent, significant predictors of the outcome variable. In this model, intervention exposure X DLL status interacted significantly to predict the
outcome variable \( (p < .001) \), whereas the intervention exposure X maternal education interaction did not reach significance. Figure 2 depicts the nature of the interaction, illustrating that exposure had a larger effect for DLL children’s score on the Language Composite compared to non-DLL children.

**Discussion**

This effectiveness trial was designed to test the impacts of a systematic, explicit language and literacy intervention implemented within real-world circumstances with an unselected, heterogeneous sample of Danish preschoolers. In contrast to prior efficacy studies conducted largely within the U.S., this study was implemented via the adoption of the intervention by country representative municipalities and under routine practice with no implementation reinforcement. We were therefore able to examine the effectiveness of two commonly used intervention components (grouping and soft-scripted lessons with aligned activities) implemented at-scale to identify which were most effective. Furthermore, all children in the randomly selected daycares were enrolled in the study, thus providing a socioeconomically diverse group of children to estimate intervention effects. Given the large sample size, it was possible to investigate intervention effects for all children within these daycare settings, while also exploring whether certain groups of children, as a function of family and child characteristics, may derive particular benefit. Finally, by tracking implementation fidelity of the intervention at the level of the individual child, we were able to address if child exposure predicted outcomes generally and for different groups of children.

Consistent with prior investigations of the effectiveness of systematic and explicit language and literacy interventions in the U.S. (e.g. Landry et al., 2009; Neuman, Newman, & Dweyer, 2011; Wasik & Hindman, 2011) and in Denmark (Authors et al., 2014, under review), this study found that effects can still be evident when such interventions are implemented in routine practice, albeit with reduced effects as compared to earlier efficacy-based research. All
LEAP variations had significant impact on two targeted pre-literacy skills, phonological awareness and letter knowledge, with effect sizes ranging from 0.18 to 0.38, corresponding to small to medium effects. However, effects on children’s language skills, namely vocabulary and comprehension, are seen only for children in LEAP-OPEN classrooms as compared to the BAU classrooms ($d = 0.33$ and $0.21$, respectively). Indeed, comparisons across all planned contrasts for both pre-literacy and language outcomes demonstrated that LEAP-OPEN was the most impactful of the LEAP variations. This is the most striking finding of the present study. It suggests that, within the current context, providing a theoretically based and empirically supported scope and sequence, without use of scripting, promotes larger gains in children than if the same scope and sequence is implemented with soft-scripted lessons. Manipulating group format, on the other hand (LEAP-LARGE vs. LEAP-SMALL), did not differentially impact child outcomes. This study therefore adds further evidence to the results of the previously mentioned meta-analysis (Marulis & Neuman, 2010) which indicate that group size is not associated with child outcomes in vocabulary whereas pre-literacy skills can be improved in large-group settings, even when taken to scale.

Moderation analyses of children participating in LEAP-OPEN indicated that only age predicted outcomes, that is, older children benefited more from the intervention than younger children whereas no other differential effects for children depending on mother’s education or ethnic background were found. Additionally, children’s exposure to the LEAP-OPEN intervention predicted outcomes, and LEAP-OPEN had a larger effect on children with a DLL background compared to children with a non-DLL background. In the following, we explore and interpret the main findings.

Scale-up, fidelity, and impact of interventions

As referred to earlier, interventions that are appropriate for scale-up are those that exhibit two core characteristics: (1) positive effects on the outcome of interest evidenced in efficacy stud-
ies, and (2) they can be implemented with fidelity (O’Donnell, 2008). This study adds to the documentation that systematic, explicit language and literacy interventions implemented at-scale do maintain impact of practical significance (although diluted to some extent) even though the implementation fidelity was reduced. Analyses of submitted implementation logs indicated that in average 75% of the 40 lessons were carried out in the two small group conditions (LEAP-SMALL and LEAP-OPEN) whereas almost 90% of the intervention was carried out in LEAP-LARGE. Educators delivered the suggested dosage in large group settings but somewhat less in small group settings. Substantial individual variation could be observed and future research must address factors that may increase fidelity. Associations between child exposure and outcomes suggest that the lower fidelity is related to the lower outcomes as suggested by earlier research (Authors, 2014, under review).

However, interesting nuances to these findings can be observed. In LEAP-SMALL and LEAP-LARGE, as seen in the other effectiveness and scaling up studies referred to above as well, the positive effects were restricted to children’s literacy outcomes. This differentiated effect is reflected in efficacy studies as well, in which effect sizes are also higher for literacy outcomes compared to oral language outcomes (0.87 vs. 0.25 respectively in NELP, 2008). It is therefore a general finding that oral domain abilities are more difficult to change than the pre-literacy domain. The fact that positive outcomes in oral language skills are generally not found in effectiveness studies warrants more attention. However, LEAP-OPEN stands out from this general picture as positive impacts were found for both oral language as well as pre-literacy outcomes. Thus, it appears that there is less of a reduction in effect for LEAP-OPEN than is seen in previous studies of systematic, explicit intervention.

**LEAP-OPEN and the role of scope and sequence versus prescribed activities**

Systematic explicit interventions are often based on a curriculum that offers well-designed lesson plans of activities, based on an understanding of children’s trajectories of learning within
specific content areas (Duncan & Magnuson, 2013). Sequencing activities according to learning trajectories and targets is thought to guide learning to be more effective and efficient and to avoid fragmentation (Clements, 2007) whereas scope and sequence is thought to ensure learning focused interactions. However, the positive results of LEAP-OPEN question whether the instructional techniques and activities work this way in practice. The comparison among the three LEAP programs demonstrated that providing a theoretically based and empirical supported scope and sequence embedded in educator-developed activities (as was the case in LEAP-OPEN) promoted larger gains in children in both pre-literacy and language domains than if the same scope and sequence was implemented with aligned activities. This is supported by the fact that lower fidelity of exposing children to the scope and sequence of learning objectives was reported by educators in LEAP-SMALL and LEAP-Large (62-65%) compared to educators in LEAP-OPEN (75%). These differences hint that educators can follow a scope and sequence to a higher extent without having soft-scripted lessons, with aligned activities perhaps because educators can focus on the most important part of the intervention -- engaging in learning focused interactions -- when they are not asked to follow a soft-scripted aligned activities. Higher levels of educator freedom in planning and implementation compared to the two soft-scripted conditions may have reduced the complexity of intervention, thereby enabling educators to focus on the need of individual children. Similarly, educator-designed activities may have made it more feasible to exploit current child engagement in specific topics for learning as indicated by the higher reported child engagement compared to the other LEAP-OPEN.

Contextual factors may also have contributed to the enhanced outcomes for LEAP-OPEN. The early educational context in Denmark has a strong focus on child-oriented practices and manu- nalized programs targeting academic achievements are uncommon (Baumüller, Gørtz & Rasmussen, 2014). Freedom of choice in terms of educational practice and methods is highly valued in the entire Danish school system (Henriksen, 2012). We can therefore speculate that the educators
in LEAP-OPEN have experienced less reluctance and higher motivation to engage in the LEAP-OPEN program compared to educators in the manualized conditions as the implementing of the program is educator-driven. These speculations have been confirmed anecdotally in focus interview with educators conducted after the LEAP study was completed but need to be investigated further. The educational level of Danish educators is rather high - and higher than in the US - as 60% of educators have a 3.5 year pedagogical bachelor degree. The educational level in Danish educators may therefore have served as a necessary background for the educators to implement and exploit the sequence and scoping to create learning focused interactions that advanced children’s language and literacy development. The generalizability of the findings should therefore be exploited in new studies.

**Child characteristics and intervention effectiveness**

Both developed and developing societies are experiencing an increasing demand for advanced literacy, academic, and cognitive skills; and improvement of early language and literacy skills predicts later educational attainment in school (NELP, 2008). Interventions that are appropriate for scale-up should therefore exhibit positive effects for children regardless of background characteristics and not only for children who are disadvantaged. We therefore examined the extent to which child characteristics including age, gender, prior language skills, SES and language status moderated the LEAP impacts and with one exception, we did not observe any differential effects (Buysse, Peisner-Feinberg, Páez, Hammer, & Knowles, 2014; Marulis & Neuman, 2013; NELP, 2008). This finding contrasts with previous cross-study conclusions, which have indicated that child characteristics are associated with risk did interact with intervention effects. In agreement with the findings from Authors (under review) we found that both children at-risk and children without any risk gained from the intervention when exposed to the same intervention. This is not a trivial finding as systematic, explicit instruction typically has been designed with skill reme-
diation rather than skill acceleration in mind. Surprisingly, the only differential effect of the intervention conditions on child outcomes were that older children gained more from the intervention compared to younger children, specifically for the pre-literacy outcome. This result may reflect that the educators have been less sensitive to the younger children’s needs.

Interestingly, we did find differential effects in terms of associations between exposure and outcomes, which indicated that DLL children who were exposed more to the intervention, gained more. There may be several reasons for this result. DLL children in Denmark underperform compared to non-DLL children, and fully one-half of DLL children with a non-western background perform more than one standard deviation below the mean of monolingual children at school start (Bleses, Højen, Jørgensen, Jensen, & Vach, 2010). This may reflect the amount and the quality (e.g. richness and variability) of exposure to Danish they have had, including that provided in Danish daycares. A study of structural and process quality in Danish daycares indicates that the proportion of non-Danish children in a classroom is negatively related to process quality (Slot, Bleses, Justice, Markussen-Brown & Højen, submitted). The current child-oriented educational practice in daycares promotes a view of less successful children like children with DLL background to have different but not deficient language skills. Low Danish language levels of DLL children have therefore been accepted as natural and have only to a limited extent caused specific targeted interventions in these children. A very straightforward explanation for the higher association between exposure and outcome in DLL children may be that the sequence and scope in LEAP-OPEN ensured multiple exposure to relevant age-appropriate language targets and more opportunities to practice Danish, but more research is needed.

Limitations

Several important limitations of this study warrant note. First, attrition was substantial, as is commonly seen in effectiveness studies. It is unclear whether similar results would have been observed if less attrition had occurred. Second, some of the subscales exhibited ceiling effects at
pretest, such that a percentage of children had no possibility of showing improvement on post-
tests. This may have led to an underestimation of effects. Third, the assessments of children’s lan-
guage and pre-literacy skills were administered by their daycare educators, and not trained re-
search staff. Independent assessments of children would have been preferable albeit difficult to
implement given the size of the sample. Finally, long-term follow-up data is not available yet;
thus it is not clear whether the obtained effects will result in improved reading skills for the in-
volved children.

Overall, the study attempted to unpack systematic, explicit language and literacy interven-
tions by investigating the effect of two components typically embedded, that is group size and
scripted lessons with aligned activities when implemented at-scale in a diverse group of children.
What seems to be most effective, at least in a Danish context, is providing educators with an ex-
plicit sequence and scope of language and literacy domains and specific learning objectives cou-
pled with requirements to conduct weekly activities. Further investigations to determine whether
this result holds with other interventions applied within early-education settings, including out-
side of the Danish context, are an important direction for future research.
## APPENDIX

Scope and sequence of instruction in English (“RIA”) and in Danish (LEAP)

### Instructional Domain 1: Vocabulary Objectives

1. To understand and use words for the names of unfamiliar objects (nouns) and actions (verbs) and that describe things and actions (adjectives and adverbs)
   1. At forstå og bruge mindre typiske navneord (ord for ting), udsagnsord (ord for handlinger), og tillægsord (ord der beskriver ting)
2. To understand and use new words representing spatial concepts (e.g., over, under, above)
   2. At forstå og bruge nye ord for rumlige begreber (fx over, under, på)
3. To understand and use new words representing time concepts (e.g., before, after, then)
   3. At forstå og bruge nye ord der har med tid at gøre (fx først, bagefter, så)
4. To talk about the meaning of new words, including how words can have more than one meaning
   4. At kunne tale om nye ords betydning (fx hvordan nogle ord kan have mere end en betydning)
5. To understand and use new words representing feelings (e.g., embarrassed, sad, joyful)
   5. At forstå og bruge nye ord der har med følelser at gøre (fx flov, bekymret, glad)
6. To understand and use new words representing thinking processes (e.g., believe, imagine)
   6. At forstå og bruge nye ord for tankeprocesser (fx tro, forestille sig)

### Instructional Domain 2: Narrative Objectives

1. To identify and describe the setting and characters of a story
   7. At kunne finde og beskrive personer og omgivelser i en historie
2. To identify and describe one or more major action(s) or event(s) in a story
   8. At kunne finde og beskrive en eller flere vigtige begivenheder i en historie
3. To order three or more major events in a story
   9. At kunne sætte mindst tre begivenheder i en historie i den tidsmæssige rigtige rækkefølge
4. To produce a fictional story that has a setting and characters
   10. At kunne fortælle en historie selv med personer og omgivelser
5. To produce a fictional or personal story that has a clear beginning, middle, and end
   11. At kunne fortælle en historie selv med en klar start, midte og slutning
6. To share feelings, ideas, or experiences in a single story that is precise and understandable
   12. At kunne fortælle om følelser, tanker, eller oplevelser i en historie som er præcis og forståelig

### Instructional Domain 3: Print Knowledge Objectives

1. To recognize that print carries meaning and to distinguish print from pictures
   1. At forstå at skrevne ord bærer betydning og forstå forskellen på billeder og ord
2. To recognize the left-to-right and top-to-bottom directionality of print
   2. At forstå at man på dansk læser fra venstre mod højre og oppe fra og ned
3. To recognize that print carries meaning and to distinguish print from pictures
   3. At kunne benævne nogle store bogstaver, fx i eget navn
3. To identify some uppercase letters, including those in own names and those of some friends or family members
4. To understand and use new words describing aspects of books (e.g., illustrator, author, cover, title page) and print (e.g., word, letter, spell, read, write)
5. To recognize the difference between letters and words
6. To recognize some common sight words, including environmental print

### Instructional Domain 4: Phonological Awareness Objectives

<table>
<thead>
<tr>
<th>Objective</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To identify when two words or sounds are the same (e.g., dog–dog, b–b) and when they are different (e.g., dog–man, d–m)</td>
<td>1. At kunne høre når to ord rimer</td>
</tr>
<tr>
<td>2. To identify when two words share a rhyming pattern</td>
<td>2. At kunne opdele ord i mindre dele, og sætte ord sammen af mindre dele</td>
</tr>
<tr>
<td>3. To produce words that share a rhyming pattern</td>
<td>3. At kunne høre når to ord begynder med samme lyd</td>
</tr>
<tr>
<td>4. To segment words into syllables and to blend syllables into words</td>
<td>4. At kunne finde på ord der rimer</td>
</tr>
<tr>
<td>5. To identify when two words share the same first sound</td>
<td>5. At kunne finde og sige et ord der begynder med en bestemt lyd</td>
</tr>
<tr>
<td>6. To produce a word starting with a specific first sound</td>
<td></td>
</tr>
</tbody>
</table>

Note: Objectives reprinted with permission of the authors of *Read It Again-PreK!* (Justice & McGinty, 2012).
References


Authors (under review). The Effectiveness of a Large-Scale Language and Pre-literacy Intervention: The SPELL Randomized-Controlled-Trial in Denmark.


Table 1  
Baseline Characteristics of Recruited Children, Daycare Centers, Classrooms, Daycare Educators, and Groups in Four Conditions

<table>
<thead>
<tr>
<th></th>
<th>BAU</th>
<th>LEAP-LARGE</th>
<th>LEAP-SMALL</th>
<th>LEAP-OPEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children, n</td>
<td>1,711</td>
<td>1,568</td>
<td>1,726</td>
<td>1,701</td>
</tr>
<tr>
<td><strong>Daycare characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daycare, n</td>
<td>40</td>
<td>36</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>Classrooms per daycare, mean</td>
<td>2.7</td>
<td>2.6</td>
<td>2.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Educators per classroom, mean</td>
<td>1.1 (3)</td>
<td>1.74 (84)</td>
<td>1.86 (63)</td>
<td></td>
</tr>
<tr>
<td>Groups per educator, mean</td>
<td>1.1</td>
<td>1.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Children per group, mean</td>
<td>18.5</td>
<td>5.7</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td><strong>Educator characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% &gt; 10 yr. teaching experience</td>
<td>61</td>
<td>67</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>% BA or higher education</td>
<td>79</td>
<td>59</td>
<td>78</td>
<td>74</td>
</tr>
<tr>
<td><strong>Parent characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal educationb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Low</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>% Low-Mid</td>
<td>33</td>
<td>34</td>
<td>41</td>
<td>36</td>
</tr>
<tr>
<td>% High-Mid</td>
<td>30</td>
<td>30</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>% High</td>
<td>24</td>
<td>24</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>Family income as mean quintile (SD)</td>
<td>3.1 (1.5)</td>
<td>3.2 (1.5)</td>
<td>3.0 (1.4)</td>
<td>3.2 (1.4)</td>
</tr>
<tr>
<td>% DLL background</td>
<td>9.5</td>
<td>8.8</td>
<td>7.7</td>
<td>7.7</td>
</tr>
<tr>
<td><strong>Child Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% boy</td>
<td>53</td>
<td>52</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>Age in months (SD)</td>
<td>54 (10)</td>
<td>53 (11)</td>
<td>53 (11)</td>
<td>53 (10)</td>
</tr>
<tr>
<td><strong>Time-1 assessments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-literacy Composite</td>
<td>-0.08 (.85)</td>
<td>.05 (.86)</td>
<td>-0.08 (.88)</td>
<td>-0.06 (.88)</td>
</tr>
<tr>
<td>Sound Discriminationc</td>
<td>14.8 (2.2)</td>
<td>14.8 (2.2)</td>
<td>14.5 (2.5)</td>
<td>14.7 (2.2)</td>
</tr>
<tr>
<td>Rhyme Detectiond</td>
<td>10.2 (3.9)</td>
<td>10.6 (4.0)</td>
<td>10.2 (4.1)</td>
<td>10.4 (4.1)</td>
</tr>
<tr>
<td>Deletiond</td>
<td>3.4 (5.0)</td>
<td>4.3 (5.7)</td>
<td>3.4 (5.0)</td>
<td>3.9 (5.6)</td>
</tr>
<tr>
<td>Letter Identificationd</td>
<td>6.8 (3.6)</td>
<td>7.3 (3.6)</td>
<td>6.5 (3.7)</td>
<td>6.7 (3.7)</td>
</tr>
<tr>
<td>Language Composite</td>
<td>-10 (.83)</td>
<td>-.04 (.86)</td>
<td>-.13 (.81)</td>
<td>-.05 (.81)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>42.0 (19.0)</td>
<td>42.6 (20.0)</td>
<td>42.6 (19.6)</td>
<td>42.6 (19.4)</td>
</tr>
<tr>
<td>Communicatione</td>
<td>29.5 (6.4)</td>
<td>28.6 (6.7)</td>
<td>28.8 (6.3)</td>
<td>29.3 (6.3)</td>
</tr>
<tr>
<td>Comprehensione</td>
<td>17.6 (4.2)</td>
<td>18.0 (4.5)</td>
<td>17.5 (4.4)</td>
<td>17.8 (4.3)</td>
</tr>
</tbody>
</table>

Note. a The unevenness arose because one daycare had two campuses which registered in the project management system separately, thereby adding an extra daycare to the BAU condition. In addition, one daycare assigned to LEAP-LARGE conducted the LEAP-SMALL program due to an administrative error. b Cf. 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%. c Administered only to 3-year-olds; d Administered only to 4-6-year-olds; e Administered to all children.
Table 2

Implementation Fidelity and Adherence to Intervention Based on Self-Report by Condition

<table>
<thead>
<tr>
<th></th>
<th>LEAP-LARGE</th>
<th>LEAP-SMALL</th>
<th>LEAP-OPEN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implementation fidelity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed lessons (max 40 lessons)</td>
<td>35.3 (7.6)</td>
<td>29.8 (9.9)</td>
<td>28.9 (9.5)</td>
</tr>
<tr>
<td>Child exposure (max 40 lessons)</td>
<td>29.3 (9.2)</td>
<td>25.8 (7.6)</td>
<td>25.2 (9.2)</td>
</tr>
<tr>
<td><strong>Adherence to core intervention elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of materials and activities (100%)</td>
<td>78.0 (.18)</td>
<td>79.0 (.23)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Children’s general performance and engagement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General performance (max 3)</td>
<td></td>
<td>2,3</td>
<td>2,6</td>
</tr>
<tr>
<td>Engagement (max 3)</td>
<td></td>
<td>2,4</td>
<td>2,7</td>
</tr>
</tbody>
</table>
Table 3
Effect-Size Estimates for Three LEAP Conditions Compared to BAU

<table>
<thead>
<tr>
<th>Variable (BAU n)</th>
<th>LEAP-LARGE</th>
<th>LEAP-SMALL</th>
<th>LEAP-OPEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sequence and scope</td>
<td>Sequence and scope</td>
<td>Sequence and scope</td>
</tr>
<tr>
<td></td>
<td>Large group</td>
<td>Small group</td>
<td>Small group</td>
</tr>
<tr>
<td>Pre-literacy Composite</td>
<td>.20*</td>
<td>.18*</td>
<td>.32**</td>
</tr>
<tr>
<td>(n = 1,419)</td>
<td>(n = 1,209)</td>
<td>(n = 1,350)</td>
<td>(n = 1,373)</td>
</tr>
<tr>
<td>Sound Discrimination</td>
<td>.09</td>
<td>-.04</td>
<td>.11</td>
</tr>
<tr>
<td>(n = 446)</td>
<td>(n = 384)</td>
<td>(n = 434)</td>
<td>(n = 467)</td>
</tr>
<tr>
<td>Rhyme Detection</td>
<td>.18*</td>
<td>.27**</td>
<td>.38**</td>
</tr>
<tr>
<td>(n = 793)</td>
<td>(n = 703)</td>
<td>(n = 765)</td>
<td>(n = 761)</td>
</tr>
<tr>
<td>Deletion</td>
<td>.16</td>
<td>.24**</td>
<td>.30**</td>
</tr>
<tr>
<td>(n = 949)</td>
<td>(n = 817)</td>
<td>(n = 891)</td>
<td>(n = 884)</td>
</tr>
<tr>
<td>Letter Identification</td>
<td>.25**</td>
<td>.25**</td>
<td>.31**</td>
</tr>
<tr>
<td>(n = 918)</td>
<td>(n = 754)</td>
<td>(n = 882)</td>
<td>(n = 830)</td>
</tr>
<tr>
<td>Language Composite</td>
<td>-.11</td>
<td>-.08</td>
<td>.30**</td>
</tr>
<tr>
<td>(n = 1,422)</td>
<td>(n = 1,213)</td>
<td>(n = 1,352)</td>
<td>(n = 1,379)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-.07</td>
<td>-.01</td>
<td>.33**</td>
</tr>
<tr>
<td>(n = 1,428)</td>
<td>(n = 1,217)</td>
<td>(n = 1,361)</td>
<td>(n = 1,382)</td>
</tr>
<tr>
<td>Comprehension</td>
<td>-.02</td>
<td>.03</td>
<td>.21**</td>
</tr>
<tr>
<td>(n = 1,417)</td>
<td>(n = 1,208)</td>
<td>(n = 1,352)</td>
<td>(n = 1,365)</td>
</tr>
<tr>
<td>Communication</td>
<td>-.11</td>
<td>-.18</td>
<td>.02</td>
</tr>
<tr>
<td>(n = 1,365)</td>
<td>(n = 1,192)</td>
<td>(n = 1,340)</td>
<td>(n = 1,313)</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01
Table 4

HLM Results for Predicting Pre-literacy and Language Composite Outcomes in the LEAP-OPEN condition from Child-Level Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Pre-literacy composite</th>
<th>Language composite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td><strong>Base model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Gender (boy)</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Pretest</td>
<td>0.47</td>
<td>0.02</td>
</tr>
<tr>
<td>Mothers Education(^1)</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>DLL status</td>
<td>-0.20</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Intervention model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEAP OPEN X Age</td>
<td>0.27</td>
<td>0.06</td>
</tr>
<tr>
<td>LEAP OPEN X Gender</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>LEAP OPEN X Pretest</td>
<td>-0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>LEAP OPEN X Mothers Education(^1)</td>
<td>-0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>LEAP OPEN X DLL</td>
<td>0.04</td>
<td>0.10</td>
</tr>
</tbody>
</table>

\(^1\)Based only on non-DLL children.
Table 5
HLM Results for Predicting Children’s Outcomes from Child Exposure (for LEAP OPEN)

<table>
<thead>
<tr>
<th></th>
<th>Pre-literacy composite</th>
<th>Language composite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Base model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Gender (boy)</td>
<td>0.18</td>
<td>0.34</td>
</tr>
<tr>
<td>Pretest</td>
<td>4.26</td>
<td>0.22</td>
</tr>
<tr>
<td>Mothers Education¹</td>
<td>0.44</td>
<td>0.22</td>
</tr>
<tr>
<td>DLL status</td>
<td>-1.72</td>
<td>0.70</td>
</tr>
<tr>
<td>Intervention Exposure</td>
<td>0.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Intervention model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interv. Exposure X Maternal Edu.¹</td>
<td>-0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>Interv. Exposure X DLL</td>
<td>0.04</td>
<td>0.07</td>
</tr>
</tbody>
</table>

¹Based only on non-DLL children.
Figure 1. Posttest scores (and 95% CI) for BAU children and children in the LEAP-OPEN condition as a function of age for the Pre-literacy Composite score.
Figure 2. Posttest scores (and 95% CI) for non-DLL and DLL children in the LEAP-OPEN condition as a function treatment exposure (mean number of lessons per child) for the Language Composite score.