

Research Article

The Impact of Interactive Shared Book Reading on Children's Language Skills: A Randomized Controlled Trial

Claire Noble,^a Thea Cameron-Faulkner,^b Andrew Jessop,^c Anna Coates,^a Hannah Sawyer,^a Rachel Taylor-Ims,^a and Caroline F. Rowland^{c,d} 

Purpose: Research has indicated that interactive shared book reading can support a wide range of early language skills and that children who are read to regularly in the early years learn language faster, enter school with a larger vocabulary, and become more successful readers at school. Despite the large volume of research suggesting interactive shared reading is beneficial for language development, two fundamental issues remain outstanding: whether shared book reading interventions are equally effective (a) for children from all socioeconomic backgrounds and (b) for a range of language skills.

Method: To address these issues, we conducted a randomized controlled trial to investigate the effects of two 6-week interactive shared reading interventions on a range of language skills in children across the socioeconomic spectrum. One hundred and fifty children aged between 2;6 and 3;0 (years;months) were randomly assigned to one

of three conditions: a pause reading, a dialogic reading, or an active shared reading control condition.

Results: The findings indicated that the interventions were effective at changing caregiver reading behaviors. However, the interventions did not boost children's language skills over and above the effect of an active reading control condition. There were also no effects of socioeconomic status.

Conclusion: This randomized controlled trial showed that caregivers from all socioeconomic backgrounds successfully adopted an interactive shared reading style. However, while the interventions were effective at increasing caregivers' use of interactive shared book reading behaviors, this did not have a significant impact on the children's language skills. The findings are discussed in terms of practical implications and future research.

Supplemental Material: <https://doi.org/10.23641/asha.12420539>

Rates of speech and/or language impairment in the United Kingdom are reported to vary between 5% and 10% (Boyle et al., 1996; Norbury et al., 2016). However, these rates are not equally distributed across the socioeconomic spectrum, with higher rates for children from disadvantaged backgrounds. This social gradient begins early, with children from different socioeconomic

backgrounds showing differences in their language processing abilities and vocabulary size from as early as 18 months of age (e.g., Fernald et al., 2013; McGillion et al., 2017). This difference in language ability appears to continue throughout the preschool and primary school years. Locke et al. (2002) reported that more than half of the children in the lowest Index of Multiple Deprivation in England started nursery school with delayed language skills, despite their general cognitive abilities being in the average range for their age. Similarly, Waldfogel and Washbrook (2010) reported that children from low-income households are, on average, 16 months behind their peers from high-income households in terms of vocabulary size at school entry. Law et al. (2011) similarly report that nearly 40% of Scottish children aged between 5 and 12 years and living in an area of pronounced deprivation had delayed language skills. Given the high rates of language delay and the significant impact poor language skills can have on a child's life (Hoff, 2013; Pace et al., 2018), the need for language interventions that are

^aDepartment of Psychological Sciences, University of Liverpool, United Kingdom

^bSchool of Arts, Languages and Cultures, The University of Manchester, United Kingdom

^cLanguage Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

^dDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

Correspondence to Caroline F. Rowland: caroline.rowland@mpi.nl

Editor-in-Chief: Sean M. Redmond

Editor: Mary Alt

Received October 17, 2019

Revision received December 20, 2019

Accepted February 26, 2020

https://doi.org/10.1044/2020_JSLHR-19-00288

Disclosure: The authors have declared that no competing interests existed at the time of publication.

accessible and effective for all socioeconomic groups is stark.

One activity that has been shown to support children's early language development is shared book reading. Research has indicated that shared book reading can support a wide range of early language skills, including vocabulary growth (e.g., Elley, 1989; Farrant & Zubrick, 2011), narrative and conversation skills (e.g., Morrow, 1988; Reese, 1995), print awareness (e.g., Justice & Ezell, 2000, 2004), future reading ability (e.g., Bus et al., 1995), and phonological awareness (e.g., Chow et al., 2008; Lefebvre et al., 2011). There is also evidence that children who are read to regularly in the early years learn language faster, enter school with a larger vocabulary, and become more successful readers at school (Bus et al., 1995).

On the basis of this research, strong emphasis has been placed on encouraging caregivers and practitioners to read with children in the early years, and many shared book reading interventions have been created to support language development and school readiness. Shared book reading interventions typically train the caregiver and/or practitioner to read with the child using a particular style, of which the most common is "interactive reading." Interactive shared reading interventions (e.g., dialogic reading) use techniques that encourage the adult to be responsive to the child and to expose the child to language that is slightly more advanced than their current language level. Interactive reading typically involves recasts, expansions, and open-ended questions, all of which have been shown to have a positive impact on a child's language development (Baker & Nelson, 1984; Cleave et al., 2015; Farrar, 1990; Girolametto & Weitzman, 2002; Huttenlocher et al., 2010; Nelson, 1977). Despite the large volume of research suggesting interactive shared book reading is beneficial for language development, two fundamental issues remain outstanding: whether interactive shared book reading interventions are equally effective (a) for children from all socioeconomic backgrounds and (b) for a range of language skills. We will discuss each issue in turn.

There has been a particularly strong focus on using interactive shared book reading interventions to improve the language and literacy skills of children from deprived backgrounds, motivated by the higher rates of language delay for such children and the desire to close this language gap. Some individual studies have reported positive gains in language outcome variables (e.g., Chacko et al., 2018; Lonigan et al., 1999; Purpura et al., 2017; Valdez-Menchaca & Whitehurst, 1992). These studies tend to train an adult (e.g., a parent, teacher, or volunteer) to read using a particular reading style and then assess the impact of the intervention on the child's language ability, most often their vocabulary skills. For example, Valdez-Menchaca and Whitehurst (1992) trained teachers to read using a dialogic reading style that involved asking more open-ended questions and responding to children's attempts to answer these questions. The intervention was aimed at low-income Mexican children attending day care and lasted between 6 and 7 weeks, and children's language skills were measured

with standardized tests of expressive and receptive language. Children who received the intervention scored significantly higher on measures of both expressive and receptive language and had a longer mean length of utterance (MLU) than children in the control group.

However, when meta-analytic methods have been used to synthesize across studies, it seems that socioeconomic status (SES) moderates the effects, with smaller effect sizes for children from disadvantaged backgrounds. For example, Mol et al. (2008) have reported that dialogic reading interventions have smaller effects on vocabulary outcomes for children at risk of language and literacy impairments ($d = 0.13$) than for children not at risk ($d = 0.53$), and here, risk status was determined by the income level and the maternal education level of the participants in the studies and is therefore a measure of SES. In a subsequent meta-analysis, Manz et al. (2010) found a similar pattern of results, reporting that shared book reading interventions had smaller effects on emergent literacy outcomes for children from low-income backgrounds ($d = 0.14$) than for children from middle- to high-income backgrounds ($d = 0.39$).

Given the need for interventions that close the language gap, these findings are of crucial importance. There are a number of possible, but not necessarily mutually exclusive, explanations that may account for this apparent difference in the effectiveness of interactive shared book reading interventions. Manz et al. (2010) have proposed that the cause of the smaller effect sizes for children from low-income families is due to a mismatch between the demands of an interactive shared book reading intervention and the parents' natural reading style. They have argued that there may be a big mismatch between the interactive style taught in reading interventions and the natural reading style of less educated parents, some of whom are more likely to focus on reading the text and describing pictures (cf. Hammer et al., 2005). This makes it harder for such parents to implement the training given in the intervention, a problem not experienced by parents who are more likely to use an interactive style naturally. Consequently, reading interventions may have a smaller effect on the language development of children from low-income families.

Another related factor is the familiarity of interactive shared book reading as an activity. Ethnographic work undertaken by McCarthey (1997) has suggested that literacy experiences may differ between children from middle-class and working-class homes. The middle-class families in McCarthey's sample tended to view reading as a form of entertainment. A range of reading material was available in the home, including children's books, and children were read to regularly by family members. In contrast, the working-class families in the sample tended to view reading as a means to maintain social relationships (e.g., reading letters and party invitations) and for religious purposes (e.g., learning passages from the bible). It may be that there are differing attitudes about the typical function of reading that make it harder for some families to adopt an intervention that requires one particular form of reading (i.e., interactive shared book reading) than others.

If this is the case, then differences in the effectiveness of interactive shared book reading interventions between families may be due to differences in the families' familiarity with the form of reading required by the intervention. Research has shown that imposing unfamiliar literacy practices, such as interactive shared reading, on a family is likely to be ineffective (Mooney et al., 2016) and that if parents do not feel comfortable with books or do not read for pleasure, then shared book reading between the parent and the child is less likely to become embedded in family practice, less likely to be sustained, and less likely to be enjoyed by children (Bus et al., 2000). Therefore, the difference in the effectiveness of interactive shared book reading interventions may be due to lower rates of shared book reading as an activity, which impacts enjoyment and engagement with the shared book reading intervention. This then leads to smaller effects on language development.

The second outstanding issue is whether interactive shared book reading interventions are equally effective for a range of language outcomes. Reviews that have synthesized the evidence indicate that the impact of shared book reading interventions is stronger for some language outcomes than others. For example, a What Works Clearinghouse report on dialogic reading indicated that this form of interactive reading has a positive impact on oral language skills but "no discernible effects" (p. 1) on phonological processing (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2007). In a more recent report on shared book reading, the authors again concluded that shared book reading did not support all language skills equally (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2015). The review indicated that shared book reading had "no discernible effects" (p. 2) on alphabets or reading achievement and reported mixed results for language comprehension and language development. In summary, while there is evidence that interactive shared book reading can support a range of language skills, when the research is synthesized and only the highest quality research is included, the results indicate that the efficacy of such interventions may vary depending on the specific outcome variables measured.

Furthermore, there are some language skills for which there is a much smaller evidence base. One such is grammatical development, for which there are only a handful of intervention studies, which have produced mixed findings. For example, Whitehurst et al. (1988) reported that children who had received an interactive shared book reading intervention developed more mature grammar abilities: They had a higher MLU, fewer single-word utterances, and higher frequency of phrases than children in a reading control group. Similarly, Valdez-Menchaca and Whitehurst (1992) reported that children receiving an interactive shared book reading intervention had longer MLUs and produced more syntactically complex sentences compared to children in a control group. However, there are also studies that have found no impact on grammatical development. For example, Lever and Sénéchal (2011) reported no difference

between the MLU of children who had received a dialogic reading intervention and that of children who had received an alternative treatment, namely, a phoneme awareness intervention. In summary, the evidence base for the impact of shared book reading on grammatical development is much weaker due to the small number of studies and the mixed findings.

In this study, we investigated whether interactive shared book reading supports language in children from a range of socioeconomic backgrounds. We used two established interactive shared book reading interventions, namely, dialogic reading (Whitehurst et al., 1988) and pause reading (Colmar, 2014), and measured the impact of the interventions on a range of language outcomes, including grammatical development. We chose these interactive shared book reading interventions as there is existing evidence that they are effective at boosting some language skills in some populations. For example, Colmar (2011, 2014) reported that pause reading supports expressive language development in children with language delay and in children with language delay from deprived backgrounds. Similarly, dialogic reading has been shown to support vocabulary development in children, but not equally for children from all socioeconomic backgrounds (e.g., Mol et al., 2008). However, these two interventions have never been directly compared in the same design, so we do not know whether one is more effective than the other, which is important information for practitioners choosing interventions to use. We explored whether the two interactive shared book reading interventions (a) were equally well implemented by high- and low-SES caregivers and (b) led to equal gains in language skills in high- and low-SES children. In addition, we included an analysis of the caregivers' book reading behavior and engagement with the intervention. The advantage of this study design was that it allowed us to explore whether there are SES differences in the effectiveness of two established interventions.

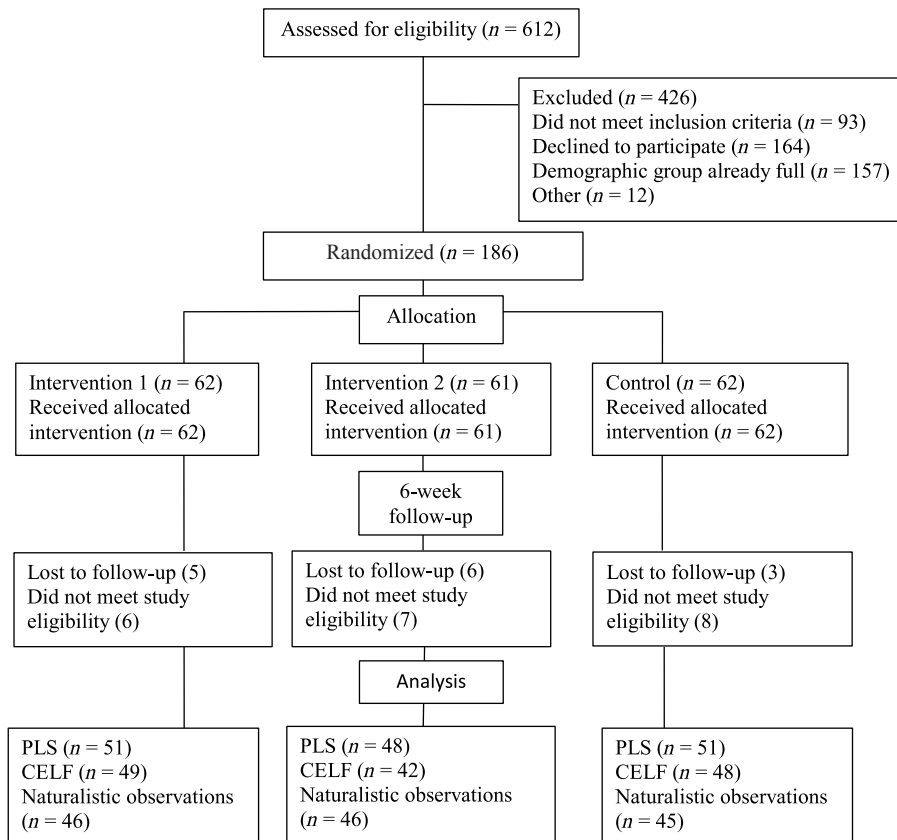
Children aged between 2;6 and 3;0 were randomly allocated to a dialogic reading intervention, a pause reading intervention, or an active reading control group. We predicted that (a) caregivers in the intervention groups would increase their interactive reading behaviors significantly more than the caregivers in the control group and (b) children in the intervention groups would have larger language gains than children in the control group. We also measured the amount of reading each dyad engaged in over the 6-week intervention.

Method

Design

This was a single-center, double-blind, parallel-group study conducted in the United Kingdom and preregistered on clinicaltrials.gov (NCT02625584). Participants were randomly assigned to one of three parallel groups in a 1:1:1 ratio, to receive one of three interventions (see Figure 1 for

Figure 1. Consolidated Standards of Reporting Trials diagram. PLS = Preschool Language Scale; CELF = Clinical Evaluation of Language Fundamentals.



the Consolidated Standards of Reporting Trials [CONSORT] diagram).

Participants

Recruitment and enrollment lasted 2 years 10 months between March 2015 and January 2018, and posttesting finished in March 2018. The project ended in April 2018 when the funding period came to an end. Eligible participants were English-speaking monolingual children aged between 2;6 and 3;0 living in North West England. Exclusion criteria were as follows: less than 37 weeks' gestation, less than 5 lbs 9 oz at birth, prolonged and/or frequent ear infections, hearing another language (not English) for more than 1 day per week, children or caregivers who had a diagnosed disability that prevented participation (e.g., inability to understand instructions). This last criterion is standard exclusion criteria in our work, but note that no families in this study were excluded on this basis. There were 150 primary caregivers, of whom 137 were mothers, 10 were fathers, two were grandmothers, and one was a childminder. Children were aged between 2;6 and 3;0 at the first visit ($M_{\text{age}} = 32$ months, $SD = 2.07$, range: 30–36), and 45% of children were female.

Caregivers' level of education ranged from no formal education to postgraduate degree. The previous literature suggests that caregiver education is the best SES predictor of language development (Arriaga et al., 1998; Bornstein et al., 1998; Dollaghan et al., 1999; Fenson et al., 1994; Hoff & Tian, 2005; Pan et al., 2005); therefore, level of caregiver education was our SES variable in all analyses.

This study received ethical approval from the University of Liverpool ethics committee. All participating caregivers gave informed consent on behalf of their child. Caregivers were reimbursed £10 for each testing session, and the child was given a book as a gift at the end of the study.

Materials

Training Videos

Videos were created to introduce the two intervention conditions and the control condition to caregivers. The intervention condition videos included clips of caregivers engaged in the target interactive reading behaviors and general advice about making reading part of daily life. The control condition video contained only information about making reading part of daily life.

Questionnaires

Caregivers completed the Family Questionnaire at the pre-intervention visit. The Family Questionnaire was devised for the UK Communicative Development Inventory project (Alcock et al., 2020) to collect information about a child's health, caregiver SES, language exposure, and whether the child attended child care. The questionnaire was constructed using a two-stage process. In the first stage, researchers created 24 questions that were hypothesized to relate to language development. In the second stage, researchers used focus groups and discussion with consultant researchers to refine and shorten the questionnaire to 21 questions. More details on questionnaire construction can be found in the work of Alcock et al. (2020). The Family Questionnaire was refined further for the current project after consultation with the original researchers to determine how successful the original version had been. Questions that were not relevant to the project were removed, and some questions were reworded to make them applicable to all family structures (see Supplemental Material S1 for the Family Questionnaire).

In the current study, we used the Family Questionnaire to (a) screen for exclusion criteria and (b) assign children to SES categories based on caregiver education. The Family Questionnaire contains seven levels of caregiver education: (1) no formal education, (2) 1–4 General Certificates of Secondary Education (GCSEs)/O-Levels (at any grade)/National Vocational Qualification (NVQ) Level 1 or similar, (3) 5+ GCSEs (Grades A*–C)/O-Levels (passes)/NVQ Level 2 or similar, (4) 1 A-Levels/2–3 AS-Levels, (5) 2+ A-Levels/NVQ Level 3 or similar, (6) University degree/Higher National Diploma/Higher National Certificate/NVQ Level 4 or 5 or similar, and (7) postgraduate degree or similar (e.g., Postgraduate Certificate in Education, PhD, MA).¹ Caregivers with a degree or higher were categorized as “high SES,” and caregivers with 2+ A-Levels or fewer qualifications were categorized as “low SES.” Note that, according to the 2011 census, 63% of the U.K. population had 2+ A-Levels or fewer qualifications (Office for National Statistics, National Records of Scotland, & Northern Ireland Statistics and Research Agency, 2016).

Caregivers also completed two further questionnaires at the pre- and post-intervention visits: the Homelife Questionnaire, which collected information on book reading in the family home, and the Title and Author Checklist, which collected information on the child's current level of exposure to storybooks. These data were not analyzed in this study.

Intervention Materials

All caregivers were given a set of 20 books to read to the child during the 6-week intervention (see Supplemental Material S2 for the list of books). All books chosen were

appropriate for the age group and were similar in length. The caregivers were also given an audio recorder and a reading diary to collect information about the amount of reading they managed over the 6-week intervention. The reading diary contained a page for each scheduled reading session, with information for the caregivers and space for comments (see Supplemental Material S3 for an example reading diary page). Each child had a different random order of books to read during the 6-week intervention period, and caregivers were asked to stick to the order in their reading diary. By following the order in the reading diary, each of the 20 books was read 3 times over the 6-week intervention period. Caregivers were instructed to audio-record the reading sessions and to mark each page in the reading diary once the book had been read.

Dialogic Reading Intervention

Caregivers were trained to read using an interactive dialogic reading style as devised by Whitehurst et al. (1988; see Box 1 in Supplemental Material S4 for a full description of the technique). Dialogic reading involves the use of a series of conversational strategies to scaffold an interactive conversation between the child and the caregiver while reading. These strategies are known as the PEER sequence. By following this sequence, the adult

- **prompts** the child to say something about the book,
- **evaluates** the child's response,
- **expands** the child's response, and
- **repeats** the prompt to help the child learn from the expansion.

A fundamental element of dialogic reading is the use of prompts to begin the PEER sequence while reading with a child. The acronym CROWD stands for five recommended prompts, as follows:

1. **Completion**
2. **Recall**
3. **Open question**
4. **Wh-question**
5. **Distancing question**

Caregivers watched a training video that explained dialogic reading in detail. Each step in the PEER sequence and each CROWD prompt were described, and video clips of a caregiver and a child reading were shown to demonstrate each step. After watching the training video, the researcher answered questions and gave the caregiver a set of information sheets that summarized the technique (see Supplemental Material S5 for the dialogic reading caregiver information sheet).

Pause Reading Intervention

Caregivers were trained to read in an interactive pause reading style based on the works of Colmar (2011, 2014; see Box 2 in Supplemental Material S6 for a full description

¹GCSEs are exams taken at the end of British high school when the student is 15 or 16 years old. Five passes at Grade C or higher are considered roughly equivalent to a U.S. high school diploma. A-Level exams are generally taken 2 years later when students are 17 or 18 years old. A-Levels are most similar to American Advanced Placement courses.

of the technique). To guide caregivers in this technique, we created the **PROB** sequence to scaffold an interactive conversation between the child and the caregiver while reading. By following the **PROB** sequence, the adult

- uses **pauses** at each page turn to let the child talk first,
- **responds** to what the child says or points to,
- uses **open-ended prompts** (i.e., asks a contingent open-ended question), and
- **boosts** what the child says by rephrasing or adding more information.

Caregivers were trained to use open-ended questions and were given the following five open-question templates to help when creating their own open questions during reading:

1. Why questions (e.g., “Why do you think...”)
2. *Tell me* prompts (e.g., “Tell me about this picture...”)
3. What questions (e.g., “What do you think...”)
4. How questions (e.g., “How do you know...”)
5. *I wonder* prompts (e.g., “I wonder what’s happening in this picture...”)

Caregivers watched a training video that explained pause reading in detail. Each step in the **PROB** sequence and each question starter were described, and video clips of a caregiver and a child reading were shown to demonstrate the technique. After watching the training video, the researcher answered questions and gave the caregiver a set of information sheets that summarized the technique (see Supplemental Material S7 for the pause reading caregiver information sheet).

Active Reading Control Condition

Caregivers in the control condition were not trained to read with their children in any specific style. Instead, these caregivers were given the same books as the children in the intervention conditions but were given only general information about how to make reading part of their daily routine. We chose to provide this information to caregivers in all conditions to help them successfully read with their children over the 6-week period.

The training video explained that reading together supports language development and provided the caregiver with the following tips to make shared reading part of their daily routine:

- Choose a space free of distractions (e.g., television and toys)
- Create an area that is comfortable and inviting
- Sit close together so the child can see the caregiver’s face and the book
- Let the child hold the book and turn the pages if they want to

- Be flexible about what time of day you read
- Choose a time of day that suits the caregiver and the child
- If the child doesn’t want to read, just follow their lead and read when they want
- Let them sit by you if they don’t want to sit still
- Have fun

Caregivers watched a training video that used clips of a caregiver and a child reading together and photos to illustrate each tip. After watching the video, the researcher answered questions and gave the caregiver a set of information sheets that summarized the information in the video (see Supplemental Material S8 for the control group caregiver information sheet).

Outcome Measures

The **Preschool Language Scale–Fifth Edition (PLS-5 UK; Zimmerman et al., 2014)** is a comprehensive language assessment instrument that evaluates both expressive and receptive language skills via elicitation and free-play and includes measures of vocabulary, phonological awareness, social communication, and language structure. The **PLS-5 UK** has good-to-excellent stability across time (test–retest reliability) with corrected stability coefficients from .86 to .95 for the different age ranges and has excellent internal consistency with split-half reliability coefficients of .96 for the age ranges tested in this study. Scores on the **PLS-5 UK** are highly correlated with scores on the **Preschool Language Scale–Fourth Edition UK** ($r = .85$) and the **Clinical Evaluation of Language Fundamentals Preschool-2 (CELF Preschool-2; $r = .79$)**, both of which are designed to test the same or similar constructs and provide evidence of its validity. It contains two subtests: the **Auditory Comprehension subscale** and the **Expressive Communication subscale**. In this study, we used the raw score on the **Auditory Comprehension subscale** as a measure of language comprehension and the raw score on the **Expressive Communication subscale** as a measure of language production.

The **CELF Preschool-2 UK (Wiig et al., 2006)** is a standardized measure of the language knowledge of individual children. In this study, we used the **Sentence Structure subtest**, which assesses children’s comprehension of a range of simple and complex sentence structures. A sentence is read to the child, and the child chooses, from a set of pictures, which picture “goes with” that sentence. The **Sentence Structure subtest** has good stability over time (test–retest reliability; $r = .77$) and good internal consistency, with a split-half reliability coefficient of .78. The **Sentence Structure subtest** is moderately correlated with the **Sentence Structure subtest of the CELF Preschool-2** ($r = .63$), and the **CELF Preschool-2 total language score** is highly correlated with the **Preschool Language Scale–Fourth Edition** ($r = .73$), both of which are designed to assess the same or similar constructs. We used the raw score on this subtest as a measure of syntax comprehension.

Finally, we calculated child mean length of utterance in morphemes (MLU) from transcriptions of the child speech during the pre-intervention session, which were transcribed in CHAT following CHAT guidelines (MacWhinney, 2000). MLU is a measure of the morphosyntactic complexity of speech. Because of the time it takes to transcribe naturalistic data, we transcribed a random sample of participants from each condition/SES level ($N = 47$; mean number of utterances per child per session = 88.67, $SD = 33.84$, range: 8–175).

Procedure

Participants were recruited from an existing database of families interested in taking part in research, by our partners in local nurseries and children centers, and through advertisements on social media. Families attended two sessions with a member of the research team: a pre-intervention session and a post-intervention session. These visits took place at one of three locations: the family home, a community setting such as the child's nursery, or the University Language Laboratory. At the pre-intervention session, (a) caregivers completed the pre-intervention questionnaires, (b) the researcher administered the PLS-5 UK and the Sentence Structure subtest of the CELF Preschool-2 UK, and (c) the child and the caregiver were video-recorded playing with toys for 10 min and reading two books.

Once both the language assessments and the naturalistic recordings were complete, the researcher randomly assigned the family to one of the following groups: dialogic reading intervention, pause reading intervention, or an active reading control group, according to CONSORT guidelines (Schulz et al., 2010). Randomization was achieved via the following procedure. Prior to the start of the project, a randomization sequence was generated by an independent researcher, unconnected to the project, with a 1:1:1 allocation using random block sizes of 3, 6, and 9. Blocks were generated with a permuted-block design using a computerized random number generator. The allocation sequence was concealed from the researchers inside sequentially numbered, opaque, sealed and signed envelopes. The piece of paper inside the envelope that stated the condition assignment was wrapped in foil to prevent the allocation being visible from the outside of the envelope.

After completion of the language assessments and naturalistic recordings, the researcher selected the envelope bearing the participant's assigned number. The researchers then video-recorded themselves opening the envelope to document that the envelope was intact prior to allocation and to record the participant's condition assignment. The caregiver remained blind to the three conditions within the study and blind to their condition assignment.

The intervention was then introduced to the caregiver using a training video specific to the condition assignment. For the dialogic reading intervention condition and the pause reading intervention condition, caregivers were introduced to the interactive reading technique and were given general advice about making reading part of their daily

routine. Caregivers in the control condition were only given the general advice about making reading part of their daily routine.

After watching the training video, the researcher summarized the content of the video and provided the caregiver with information sheets specific to each condition. The researcher also provided links to watch the video again and to access audio recordings of the summary leaflet. The researcher explained how to use and complete the reading diary as well as how to use the audio recorder and asked the families to try to read two books to their child 5 times a week.

During the 6-week intervention period, the researchers contacted all caregivers by e-mail, text, phone call, or letter on a weekly basis. Through these weekly contacts, the intervention messages were repeated, and the caregivers' questions were answered.

Approximately 6 weeks after the pre-intervention session, the caregiver and the child attended a post-intervention session (mean number of days between pre- and post-intervention sessions = 47.66 days, $SD = 6.22$, range: 41–70). At this visit, the language assessments and the naturalistic recordings from the pre-intervention session were repeated by a researcher blind to the condition assignment of the family. The caregiver also completed the post-intervention questionnaires. At the end of the session, the caregivers were given a full debrief about the purpose of the study.

Intensity and Number of Reading Sessions

Each caregiver was asked to read 5 times per week and to read two books per session. This equaled a maximum of 60 reading sessions over a 6-week intervention period. The choice of a 6-week duration was based on previous studies that have reported positive findings of interactive reading on the language development of children in our target age range using interventions of similar length (most run between 4 and 8 weeks). For example, in Mol et al.'s (2008) meta-analysis, three out of the four reading interventions involving 2- to 3-year-old children were all between 4 and 6 weeks in duration and typically reported medium-to-large effect sizes on productive measures of language development.

The caregivers were given an audio recorder and a reading diary to collect information about their reading behaviors. Caregivers were asked to audio-record all reading sessions and to use the reading diary to document when they read. To calculate the number of reading sessions, we counted the sessions marked in the reading diary. If fewer than 60 sessions (i.e., the maximum number of sessions) were marked in the reading diary, we cross-referenced with the audio recordings to check for any additional reading sessions that were recorded but not marked in the reading diary. One hundred and forty-seven of the 150 caregivers provided this information (note that the other three were not categorized as "lost to follow-up" in the CONSORT diagram [see Figure 1] but were included in the analyses). The mean rate of reading sessions was high, but there was a large variation in the number of sessions between families ($M = 50.63$, $SD = 14.54$, range: 0–60).

Coding

The pre- and post-intervention book reading video recordings were coded for the presence of caregiver dialogic reading and pause reading behaviors. All participant video files were coded for both types of book reading behaviors, regardless of the condition the participant was assigned to, and coders were blind to the condition assignment of the participants when coding.

Most naturalistic analyses in the child language literature use an event-counting method to code behavior, in which researchers transcribe and code every utterance produced by both the child and caregivers. However, this is extremely time consuming and is often not practical or cost effective in studies, such as this one, that test a large number of participants. In addition, the data it yields are not necessarily as representative of the child's or the caregivers' everyday behavior, as is commonly assumed (see Tomasello & Stahl, 2004, for a detailed description of the problems inherent in this observational sampling method). Thus, for this study, we chose a different observational coding method, namely, a "one-zero time sampling" method, in which the observer records, during each sample period, whether the behavior occurred at least once (scored as 1) or did not occur (scored as 0). This method is commonly used in the animal behavior literature (see Martin & Bateson, 1986, for a review) and has been used for coding child behaviors in the past (albeit not as often recently; e.g., see Bishop, 1951; Olson, 1929; Richards & Bernal, 1972). Each video was split into 30-s segments, and researchers coded whether there were dialogic reading and pause reading behaviors present in each 30-s segment (see Supplemental Material S9 for details of the coding scheme). This method is suitable for our purposes (achieving a reliable estimate of differences in the target behaviors produced across conditions and across time) because the target behaviors are of variable duration yet total observation time is relatively constant across conditions. The method allows us to equate caregivers who respond differently but equally effectively to the training. The dependent variable was the proportion of the segments that contained evidence of the target book reading behaviors. This resulted in two scores of book reading behavior per video: one score for caregiver dialogic reading behaviors and one score for caregiver pause reading behaviors. A randomly selected 10% of both the pre- and post-intervention recordings were second coded. For dialogic reading behaviors, Cohen's κ was .85, and for pause reading behaviors, Cohen's κ was .91, indicating excellent agreement on both measures. All discrepancies were resolved by the first author.

Results

Analysis Strategy

The data were analyzed using SPSS (Version 22) and R (Version 3.6.0; R Core Team, 2018) using RStudio (Version 1.2.1335; RStudio Team, 2019; data, scripts, and output files are available in an R project folder at [\[osf.io/txu63/\]\(https://osf.io/txu63/\); see also Supplemental Material S13\). For all regression models, to increase the stability of the model, the predictors were not fully crossed. Instead, the model only included the interactions of theoretical interest. In practice, this meant that the interactions between caregiver reading behavior, pre-intervention score, and family SES were not included, but the interaction of each of these predictors with the intervention group was included, to establish whether the potential effects of the intervention groups were dependent on other variables. The reported statistics were generated using bootstrapped simulations \(\$R = 1,000\$ \), which provided 95% confidence intervals as well as standard errors and \$p\$ values derived from the sampling distribution. Descriptive statistics for the main analyses are provided in violin plots below and in tables \(raw score means and standard deviations\) in Supplemental Material S10.](https://</p></div><div data-bbox=)

Preliminary Analyses

We ran a series of one-way analysis of variance (ANOVA) to test for differences across conditions in the age and SES of the participants and the level of caregiver education (see Table 1). There were no significant differences in age, $F(2, 144) = 1.67, p = .19, \eta_p^2 = .02$; SES (low/high), $F(1, 144) = 0.48, p = .49, \eta_p^2 = .003$; and education, $F(2, 147) = 0.096, p = .91, \eta_p^2 = .001$.

Caregiver Reading Style

The first analysis examined whether the reading interventions had the intended effect on the caregiver's reading style by comparing the presence of dialogic and pause reading behaviors in the pre-intervention reading session and the post-intervention reading session 6 weeks later.

At the pre-intervention session, there were similar levels of dialogic and pause reading behaviors across the intervention and control groups (dialogic reading behavior: dialogic group, $M = 0.41, SD = 0.22$, and control group, $M = 0.42, SD = 0.23$; pause reading behavior: pause group, $M = 0.17, SD = 0.12$, and control group, $M = 0.25, SD = 0.17$); although, note that, overall, the parents were more likely to produce dialogic reading behavior ($M_s = 0.41$ and 0.42) than pause reading behavior ($M_s = 0.17$ and 0.25). Hence, separate difference scores were calculated between the pre- and post-intervention assessments for dialogic reading behaviors and pause reading behaviors, with larger scores representing a greater increase in rate of reading behaviors. Using difference scores, rather than raw pre- and post-intervention scores, allowed us to simplify the regression model and simulate the main comparisons of interest as main effects, for ease of interpretation.

Figures 2 (dialogic reading) and 3 (pause reading) illustrate the results. Dialogic reading scores increased in the dialogic reading group (pre-intervention: $M = 0.41, SD = 0.22$; post-intervention: $M = 0.71, SD = 0.20$) but not in the control group (pre-intervention: $M = 0.42, SD = 0.23$; post-intervention: $M = 0.39, SD = 0.25$). Similarly,

Table 1. Age in months (*SD*), caregiver education level, and number and gender of child participant at the pre-intervention session.

Condition		Age in months (<i>SD</i>)	Caregiver education	<i>N</i> (girls)
Dialogic reading	High SES	32.69 (2.03)	6.45 (0.51)	29 (15)
	Low SES	32.41 (2.24)	4.32 (1.09)	22 (16)
Pause reading	High SES	32.17 (2.22)	6.45 (0.51)	29 (11)
	Low SES	31.84 (2.01)	4.05 (1.22)	19 (7)
Control	High SES	31.87 (1.83)	6.40 (0.50)	30 (10)
	Low SES	31.76 (2.14)	4.00 (1.10)	21 (9)

pause reading scores increased in the pause reading group (pre-intervention: $M = 0.17$, $SD = 0.12$; post-intervention: $M = 0.59$, $SD = 0.28$) but not in the control group (pre-intervention: $M = 0.25$, $SD = 0.17$; post-intervention: $M = 0.24$, $SD = 0.18$); see the descriptive statistics in Supplemental Material S10 for tables of raw score means and standard deviations.

We ran two multiple regression models to examine whether there were changes in parental reading behavior between the pre- and post-intervention sessions. The first compared levels of dialogic reading in caregivers receiving a dialogic intervention to those in the control group, and the second compared the pause reading behavior of caregivers assigned to the pause reading intervention to those in the control group. The outcome measure was the difference in scores for dialogic (Analysis 1) or pause (Analysis 2) reading behavior between the pre- and post-intervention recorded book reading sessions. Intervention group (dialogic/pause vs. control) and family SES (high/low) were entered as effect-coded factors. Pre-intervention scores were added as predictors to control for baseline differences, and number of reading sessions completed (No. sessions) was added to

control for the effect of the number of reading sessions completed. These were added as centered, continuous predictor variables.

The first regression model compared dialogic reading and control groups (see Table 2). The model explained 58.66% of the variance in the change in parental dialogic reading behavior, $F(7, 82) = 16.62$ [0.97, 26.84], $p < .001$, $R^2 = .59$. As predicted, the caregivers assigned to the dialogic reading intervention showed a significantly larger increase in dialogic reading style than those in the control group, $\beta = .36$ [0.17, 0.41], $SE = 0.06$, $t = 5.97$, $p < .001$. Note also that caregivers with higher pre-intervention dialogic reading scores showed smaller increases than those who started with lower scores, $\beta = -0.73$ [-0.9, -0.56], $SE = 0.09$, $t = -8.40$, $p < .001$, suggesting that the intervention had a bigger effect on those caregivers who produced fewer dialogic reading behaviors spontaneously. None of the covariates (No. sessions, SES, pre-intervention score) interacted with the intervention group, and there were no other significant main effects in the model.

The second regression model compared the pause reading and control groups (see Table 3). This model

Figure 2. Gains in dialogic reading behaviors between pre- and post-intervention reading sessions for caregivers in the dialogic reading condition and the control condition. SES = socioeconomic status.

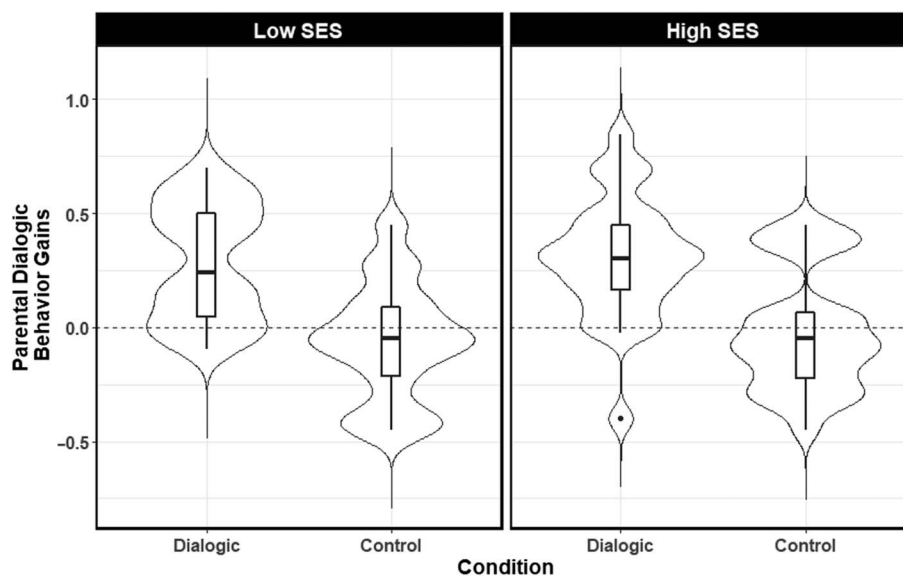
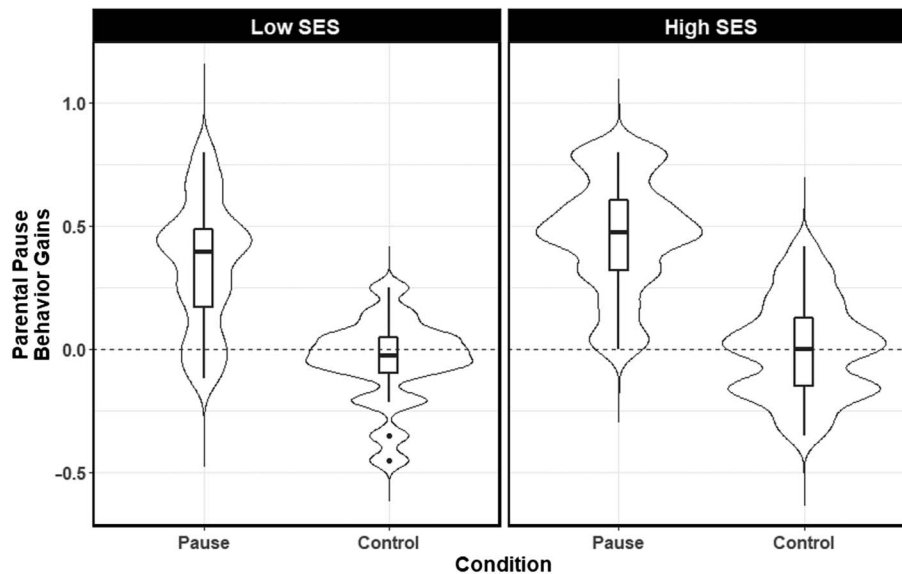


Figure 3. Gains in pause reading behaviors between pre- and post-intervention reading sessions for caregivers in the pause reading condition and the control condition. SES = socioeconomic status.



explained 58.86% of the variance in the change in parental pause reading, $F(7, 82) = 16.76 [3.96, 24.57]$, $p < .001$, $R^2 = .59$. As predicted, the caregivers assigned to the pause reading intervention showed a significantly larger increase in pause reading style than those in the control group, $\beta = .42 [0.29, 0.57]$, $SE = 0.07$, $t = 5.87$, $p < .001$. This difference interacted with the number of reading sessions completed: A greater number of sessions was associated with a larger difference in pause reading behavior change between the two groups, $\beta = 0.56 [0.05, 1.14]$, $SE = 0.28$, $t = 1.98$, $p = .047$. There was also a significant main effect of the number of reading sessions, $\beta = -0.3 [-0.58, -0.02]$, $SE = 0.14$, $t = -2.13$, $p = .034$, reflecting the larger increases in post-intervention pause reading for caregivers who had completed more sessions. There were no other significant main effects or interactions.

Collectively, the findings suggest that both interventions were effective at changing caregivers' reading style

regardless of their SES. When caregivers were trained to use either a pause reading style or a dialogic reading style, they increased their use of these reading behaviors significantly more than caregivers in the control condition.

Child Language Outcomes

Measures used were receptive language (PLS-5 UK), expressive language (PLS-5 UK), syntax comprehension (CELF Preschool-2 UK), and syntax production (MLU calculated from naturalistic data) both at pre-intervention and during a post-intervention session approximately 6 weeks later. Although pre-intervention scores for high-SES children were slightly higher, as is to be expected (see Table 4), there were similar levels of language across the intervention groups at the pre-intervention session (see Table 5). Thus, difference scores between the pre- and post-intervention assessments were calculated for each of the four language

Table 2. Summary of the multiple regression model fitted to caregiver dialogic reading behavior.

Term	β	SE	t	p
Intercept	0.12 [0.09, 0.21]	0.03	3.90	< .001
Dialogic vs. control	0.36 [0.17, 0.41]	0.06	5.97	< .001
No. sessions	2.3e-03 [-8.6e-04, 5.1e-03]	1.5e-03	1.52	.096
SES	0.03 [-0.09, 0.09]	0.04	0.76	.449
Pre-intervention score	-0.73 [-0.9, -0.56]	0.09	-8.40	< .001
Dialogic vs. control \times No. sessions	-2.6e-03 [-8.4e-03, 3.3e-03]	3.0e-03	-0.88	.304
Dialogic vs. control \times SES	-0.07 [-0.17, 0.18]	0.09	-0.76	.441
Dialogic vs. control \times Pre-intervention	-0.16 [-0.49, 0.17]	0.17	-0.91	.356

Note. This regression model explained 58.66% of the variance in the change in parental dialogic reading behavior: $F(7, 82) = 16.62 [0.97, 26.84]$, $p < .001$, $R^2 = .5866$, $N = 90$. SES = socioeconomic status.

Table 3. Summary of the multiple regression model fitted to caregiver pause reading behavior.

Term	β	SE	t	p
Intercept	0.2 [0.18, 0.32]	0.04	5.57	< .001
Pause vs. control	0.42 [0.29, 0.57]	0.07	5.87	< .001
No. sessions	-0.3 [-0.58, -0.02]	0.14	-2.13	.034
SES	0.05 [-0.09, 0.09]	0.05	1.04	.294
Pre-intervention score	7.6e-04 [-2.6e-03, 4e-03]	1.7e-03	0.45	.637
Pause vs. control \times No. sessions	0.56 [0.05, 1.14]	0.28	1.98	.047
Pause vs. control \times SES	9.7e-03 [-0.18, 0.19]	0.09	0.10	.917
Pause vs. control \times Pre-intervention	-3.6e-03 [-9.9e-03, 3.4e-03]	3.4e-03	-1.06	.267

Note. This regression model explained 58.86% of the variance in the change in parental pause reading: $F(7, 82) = 16.76 [3.96, 24.57]$, $p < .001$, $R^2 = .5886$, $N = 90$. SES = socioeconomic status.

measures, with larger scores representing greater improvement (see Figure 4).

Four separate models were fitted to these difference scores, one for each of the four language outcome measures. As is usual when building regression models in R, the factor with three levels (intervention group) was entered as a Helmert-coded factor with two contrasts. The first contrast compared the scores of the participants assigned to the pause reading group to those of the participants assigned to the dialogic reading group. The second then compared the scores of the participants receiving an intervention (i.e., the pause group and the dialogic group combined) to those of the participants in the control group. No. sessions was entered as a centered, continuous variable, based on the information collected from audio recordings and reading diaries, and pre-intervention score was included as a centered, continuous predictor. Finally, SES was entered as an effect-coded factor (high/low). The full statistics for all models, including the goodness of fit of the models to the data (R^2 values), are reported in Tables 4–7. For succinctness, we summarize only the main effects of interest in the text.

The first model was fitted to the change in expressive language scores (PLS-5 UK), with intervention group (dialogic/pause/control), SES (high/low), No. sessions, and pre-intervention score as model parameters (see Table 6). This model structure did not provide an adequate fit to the data, $F(11, 135) = 0.5 [-1.85, 0.85]$, $p = .898$, $R^2 = .04$, indicating that our predictors did not capture significant variance in the outcome measure. Contrary to our prediction, intervention group (pause vs. dialogue, control vs. intervention) did not have an effect on the change in expressive language scores between pre- and post-intervention

sessions (both $ps > .60$). No. sessions, pre-intervention score, and SES all had nonsignificant effects ($ps > .124$).

The second model was fitted to the change in receptive language scores (PLS-5 UK), which also included intervention group, SES, pre-intervention score, and No. sessions as predictors (see Table 7). Overall, the children experienced a significant increase in their receptive language over time, $\beta = 1.68 [1.7, 3.69]$, $SE = 0.51$, $t = 3.29$, $p = .002$, with the regression model, as a whole, providing a significant fit, $F(11, 135) = 1.97 [-1.53, 2.91]$, $p = .036$, $R^2 = .14$, and explaining 14% of the variance in the data. However, the only significant individual predictor in this model was pre-intervention score: The children with the largest starting vocabularies experienced less vocabulary growth between the pre- and post-intervention assessments, $\beta = -0.2 [-0.33, -0.09]$, $SE = 0.06$, $t = -3.28$, $p = .001$. This means that, contrary to our prediction, intervention group (pause vs. dialogue, control vs. intervention) did not have a significant effect on the change in receptive language scores between pre- and post-intervention sessions.

A reviewer pointed out that it might be more appropriate to use the PLS-5 Growth Scale Values (GSV scores) rather than raw scores for the two analyses above, since GSVs represent raw scores on an equal interval scale, where a one-unit increase in scores represents the same amount of change regardless of where the score falls in the developmental continuum. Thus, we converted the receptive and expressive language raw scores described above into GSV scores and reran the analyses. The full results for these analyses can be found in Supplemental Material S12 (see Sections 3.1.3. and 3.2.3). The pattern of results was the same as for the raw scores analysis. Most importantly, for both expressive and receptive vocabulary, once again,

Table 4. Pre-intervention mean (*SD*) raw scores for each of the four language measures by socioeconomic status (SES; note that mean length of utterance [MLU] scores are calculated for only a sample of the full data set).

SES	Expressive language (PLS)	Receptive language (PLS)	Syntax comprehension (CELF)	MLU
Low	34.09 (4.78)	33.56 (5.98)	4.66 (3.80)	2.56 (0.70)
High	37.40 (4.10)	37.70 (5.21)	7.08 (4.45)	2.71 (0.59)

Note. PLS = Preschool Language Scale; CELF = Clinical Evaluation of Language Fundamentals.

Table 5. Pre-intervention mean (SD) raw scores for each of the four language measures by intervention condition (note that mean length of utterance [MLU] scores are calculated for only a sample of the full data set).

Condition	Expressive language (PLS)	Receptive language (PLS)	Syntax comprehension (CELF)	MLU
Pause	35.79 (5.29)	35.48 (6.26)	5.38 (4.70)	2.58 (0.77)
Dialogic	36.96 (4.04)	36.63 (5.77)	7.02 (4.41)	2.65 (0.47)
Control	35.39 (4.53)	35.47 (5.62)	5.83 (3.91)	2.68 (0.70)

Note. PLS = Preschool Language Scale; CELF = Clinical Evaluation of Language Fundamentals.

there was no effect of intervention group (pause vs. dialogue, control vs. intervention; all $ps > .05$).

The third model was fitted to the change in participants' syntax comprehension (CELF Preschool-2 UK) scores, which also included intervention group, SES, pre-intervention score, and No. sessions as predictors (see Table 8). This regression model provided a significant fit to the data, $F(11, 124) = 2.07 [-1.34, 2.97]$, $p = .027$, $R^2 = .16$, explaining 16% of the total variance. However, the only significant predictor in the model was pre-intervention score, as the children with the highest CELF scores at the pre-intervention session showed a smaller increase in these scores over the course of the intervention, $\beta = -0.35 [-0.49, -0.21]$, $SE = 0.07$, $t = -4.75$, $p < .001$. Thus, contrary to our prediction, intervention group (pause vs. dialogue, control vs. intervention) did not have an effect on the change in syntax comprehension scores between pre- and post-intervention sessions.

In a fourth and final regression model, the participants' post-intervention increase in MLU was assessed,

with intervention group, family SES, pre-intervention score, and No. sessions as predictors (see Table 9). Note that this analysis was conducted on a random sample of participants from each condition/SES level ($N = 47$). The model did not provide a significant fit to the data, $F(11, 35) = 1.1 [-4.18, 2.51]$, $p = .388$, $R^2 = .26$, indicating that our predictors, as a whole, did not capture significant variance in the outcome measure. Most importantly, contrary to our prediction, intervention group (pause vs. dialogue, control vs. intervention) did not have an effect on the change in expressive language scores between pre- and post-intervention sessions.

To summarize, the model structures for expressive language and MLU did not provide an adequate fit to the data (R^2 values were nonsignificant). The model structures for receptive language and syntax comprehension did provide an adequate fit to the data but explained only a small amount of the total variance. SES and No. sessions were not adequate predictors in any of the models, and pre-intervention score was a significant predictor only for

Figure 4. Change in expressive language, receptive language, syntax comprehension (Clinical Evaluation of Language Fundamentals), and mean length of utterance (MLU) in each intervention condition for high- and low-socioeconomic status (SES) groups.

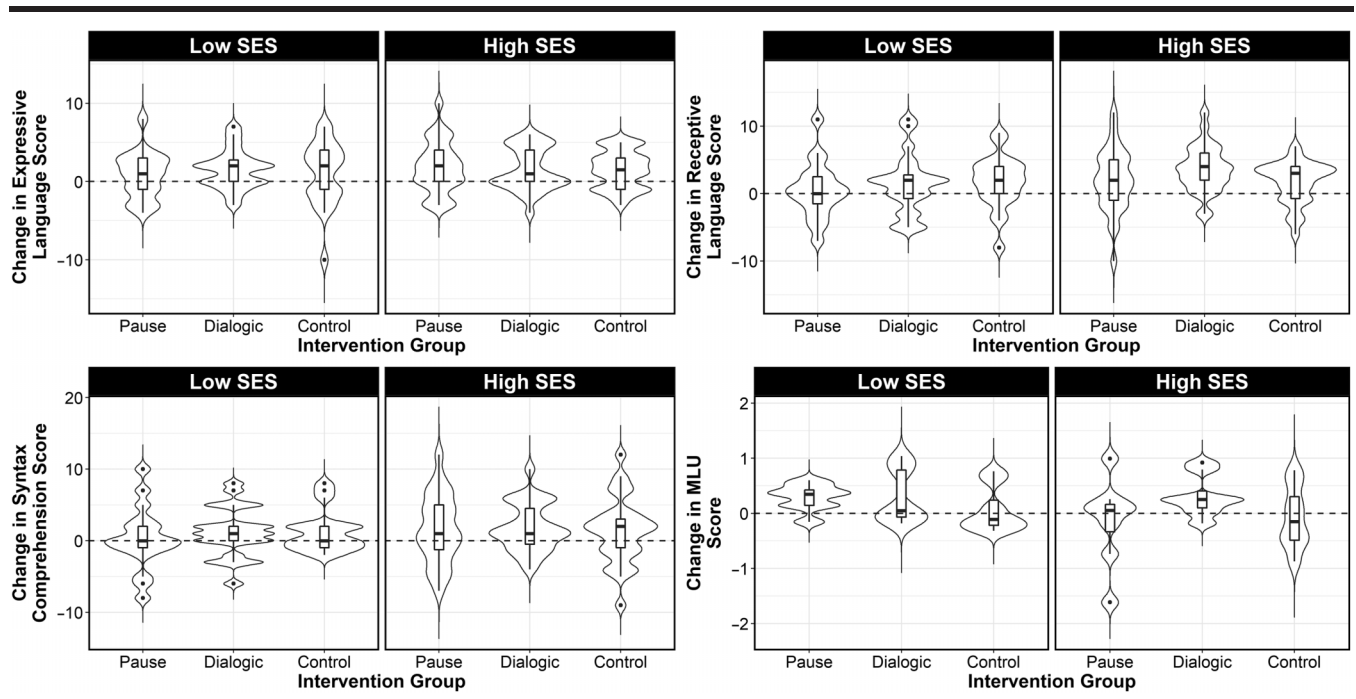


Table 6. Summary of the multiple regression model fitted to child expressive language.

Term	β	SE	t	p
Intercept	1.33 [0.83, 2.71]	0.48	2.77	.007
Pause vs. dialogic	0.5 [-1.97, 1.88]	0.98	0.51	.608
Control vs. intervention	0.24 [-1.4, 1.68]	0.79	0.31	.766
Pre-intervention score	-0.1 [-0.22, 0.03]	0.06	-1.53	.124
SES	0.42 [-1.17, 1.16]	0.59	0.71	.468
No. sessions	3.3e-03 [-0.04, 0.04]	0.02	0.16	.842
Pause vs. dialogic \times Pre-intervention Score	-0.1 [-0.41, 0.2]	0.16	-0.65	.511
Pause vs. dialogic \times SES	-0.52 [-2.59, 2.65]	1.34	-0.39	.697
Pause vs. dialogic \times No. sessions	0.03 [-0.04, 0.1]	0.04	0.82	.394
Control vs. intervention \times Pre-intervention Score	0.05 [-0.13, 0.22]	0.09	0.56	.568
Control vs. intervention \times SES	-0.03 [-1.76, 1.76]	0.90	-0.03	.973
Control vs. intervention \times No. sessions	-8.8e-03 [-0.07, 0.07]	0.03	-0.25	.738

Note. This regression model did not provide an adequate fit to the data, indicating that our predictors did not capture significant variance in the outcome measure: $F(11, 135) = 0.5 [-1.85, 0.85]$, $p = .898$, $R^2 = .0394$, $N = 147$. SES = socioeconomic status.

receptive language and syntax comprehension. Most importantly, contrary to our predictions, there was no effect of intervention group in any of the models. In other words, there were no significant differences in the gains that could be attributable to the intervention group assignment of the child (dialogic, pause, or control).²

Exploratory Analyses

Above, we concluded that there was no effect of intervention group in any of the models on any of our four language outcomes. We performed additional analyses to determine whether we can confidently state that the effects of the intervention group on language outcomes are, indeed, absent. When one runs intervention studies that yield null results, it is important to establish whether the results can, with a certain level of confidence, be attributed to a real lack of an effect or whether the most likely explanation is a statistical confound such as lack of power. Supplementary analyses are required to do this because absence of evidence is not the same as evidence of absence in the type of statistical models used in this study (frequentist models),

²In response to a reviewer's suggestion, we repeated all four analyses with the caregivers' post-intervention pause and dialogic reading scores instead of the number of reading sessions. Because these are alternative analyses of the same hypotheses, the results should be treated with caution and are reported only in Supplemental Material S11. These results also fail to support our prediction that dialogic/pause reading behavior training leads to bigger language gains. For receptive language, although the regression model was not significant overall, there was an interaction between intervention group (intervention vs. control) and reading behavior. The children in both intervention groups exposed to more pause reading behaviors had greater receptive language gains, and children in both intervention groups exposed to more dialogic reading behaviors had smaller receptive language gains. However, this did not differ by intervention type (pause vs. dialogic) and was not replicated for the other three language measures. Thus, since the analysis is post hoc and we do not see similar patterns across all four language outcome measures, we are reluctant to draw robust conclusions without replication.

since it is not "statistically or logically correct to conclude the absence of an effect when a nonsignificant effect has been observed" (Lakens et al., 2018, p. 45). Thus, in accordance with Dienes (2014) and Lakens et al. (2020), we used power analysis and equivalence tests to determine the likelihood of detecting a significant effect with the observed effect size and the collected sample size, in addition to the bootstrapped confidence intervals previously presented.

First, we performed a series of post hoc power simulations to assess whether the sample sizes recruited in this study provided sufficient power to detect effects in our data, if such effects exist. This involved resampling the data with replacement and refitting the models used in the main analysis ($R = 1,000$ simulations) and then performing further power simulations to identify the sample size that would be necessary to reach 80% power with the effect sizes we observed. The simulations were set to terminate at 10,000 participants, so this is the upper bound. Table 10 reports the power (β) levels for the main effects of interest in the regression models ("intervention vs. control" group, "pause vs. dialogic" group) at the sample sizes collected in this research. It also reports the sample sizes that would be needed for us to observe significant effects with the observed effect sizes 80% of the time.

These data suggest that our study was not underpowered (i.e., that we did have enough power to detect the predicted effects if they existed). For all contrasts reported in Table 8, the proportion of simulations that yielded a p value of less than .05 at the recruited sample sizes was low (below 26% in all cases), and for all contrasts, we would need substantially larger sample sizes for significant differences to be observed 80% of the time. For example, even for the contrast with the largest power level (between pooled intervention and control groups for $MLU = .255$), we would need a total sample size of 294 to observe a significant difference in MLU groups 80% of the time.

However, it is arguably more important to determine whether our study is powered to detect meaningful effects than it is to determine the sample size needed to detect

Table 7. Summary of the multiple regression model fitted to child receptive language.

Term	β	SE	t	p
Intercept	1.68 [1.7, 3.69]	0.51	3.29	.002
Pause vs. dialogic	1.48 [-0.69, 4.28]	1.27	1.17	.238
Control vs. intervention	0.33 [-0.57, 2.16]	0.70	0.47	.630
Pre-intervention score	-0.2 [-0.33, -0.09]	0.06	-3.28	.001
SES	1.01 [-1.23, 1.26]	0.64	1.59	.113
No. sessions	0.02 [-0.05, 0.07]	0.03	0.55	.524
Pause vs. dialogic \times Pre-intervention Score	-6.7e-03 [-0.32, 0.32]	0.16	-0.04	.967
Pause vs. dialogic \times SES	0.34 [-3.17, 3.22]	1.63	0.21	.833
Pause vs. dialogic \times No. sessions	0.04 [-0.07, 0.17]	0.06	0.74	.460
Control vs. intervention \times Pre-intervention Score	0.02 [-0.13, 0.18]	0.08	0.31	.755
Control vs. intervention \times SES	0.59 [-1.67, 1.69]	0.86	0.69	.488
Control vs. intervention \times No. sessions	-0.03 [-0.11, 0.08]	0.05	-0.60	.433

Note. This regression model provided a significant fit to the data, explaining 14% of the variance: $F(11, 135) = 1.97 [-1.53, 2.91]$, $p = .036$, $R^2 = .1382$, $N = 147$. SES = socioeconomic status.

small effects with sufficient power. Hence, for our second analysis, we ran equivalence tests using the two one-sided tests procedure (Lakens et al., 2020). Equivalence tests provide a robust way of examining whether there are no meaningful differences across the intervention and SES groups. In other words, they allow us to determine whether we can reject the presence of effects as large as, or larger than, a minimal effect size of interest and accept the null hypothesis of equivalence. For the purpose of these analyses, we specified a Cohen's d of 0.5 as the minimal effect size of interest, given the previous literature (Colmar, 2014; Manz et al., 2010; Mol et al., 2008).

Table 11 reports the results for the two main contrasts of interest ("control vs. pooled intervention" group, "pause vs. dialogic" group). The full set of results can be found in the R project folder at the clinical trials website (<https://osf.io/txu63/>). For expressive language, receptive language, and syntax comprehension, the probability of detecting the presence of effects as large as, or larger than, Cohen's $d = 0.5$ is extremely low (below 16%). For these three language outcomes, we can be reasonably confident

that we have enough power to detect/reject effects of $d = 0.5$ or higher. For MLU, the probabilities are higher (.432 and .369), but not high enough for us to detect/reject the effect confidently. Thus, for MLU, we conclude that we do not have enough power to make a confident judgment about whether effects as large as, or larger than, $d = 0.5$ exist.

Discussion

This randomized controlled trial investigated whether two interactive shared book reading interventions support a range of language skills in children from all socioeconomic backgrounds. With regard to caregiver reading behavior, we found that caregivers in the control and intervention conditions exhibited similar levels of interactive shared book reading behaviors at the pre-intervention session. However, as predicted, by the post-intervention session, caregivers assigned to the intervention groups showed a significantly larger increase in the targeted interactive shared reading behaviors than those in the control group. This

Table 8. Summary of the multiple regression model fitted to child syntax comprehension.

Term	β	SE	t	p
Intercept	1.19 [0.78, 3.07]	0.59	2.04	.041
Pause vs. dialogic	0.87 [-2.17, 3.59]	1.47	0.59	.546
Control vs. intervention	0.18 [-1.38, 1.81]	0.81	0.22	.820
Pre-intervention score	-0.35 [-0.49, -0.21]	0.07	-4.75	< .001
SES	0.63 [-1.44, 1.42]	0.73	0.86	.387
No. sessions	0.04 [-0.07, 0.12]	0.05	0.78	.445
Pause vs. dialogic \times Pre-intervention Score	0.23 [-0.11, 0.56]	0.17	1.33	.183
Pause vs. dialogic \times SES	-0.16 [-3.71, 3.81]	1.92	-0.08	.934
Pause vs. dialogic \times No. sessions	-0.02 [-0.13, 0.09]	0.05	-0.33	.722
Control vs. intervention \times Pre-intervention Score	0.06 [-0.15, 0.27]	0.11	0.60	.544
Control vs. intervention \times SES	0.2 [-1.97, 1.81]	0.96	0.21	.833
Control vs. intervention \times No. sessions	-9.0e-04 [-0.14, 0.21]	0.09	-0.01	.994

Note. This regression model provided a significant fit to the data, explaining 16% of the total variance: $F(11, 124) = 2.07 [-1.34, 2.97]$, $p = .027$, $R^2 = .1552$, $N = 136$. SES = socioeconomic status.

Table 9. Summary of the multiple regression model fitted to child mean length of utterance.

Term	β	SE	t	p
Intercept	0.12 [-0.26, 0.38]	0.16	0.75	.315
Pause vs. dialogic	0.26 [-0.2, 1.06]	0.32	0.81	.342
Control vs. intervention	0.13 [-0.44, 0.58]	0.26	0.51	.444
Pre-intervention score	-0.15 [-0.55, 0.26]	0.21	-0.74	.437
SES	-0.07 [-0.4, 0.35]	0.19	-0.38	.678
No. sessions	6.1e-03 [-0.02, 0.03]	0.01	0.44	.434
Pause vs. dialogic \times Pre-intervention Score	0.3 [-0.67, 1.22]	0.48	0.62	.494
Pause vs. dialogic \times SES	0.18 [-0.84, 0.95]	0.46	0.38	.652
Pause vs. dialogic \times No. sessions	8.8e-03 [-0.04, 0.06]	0.02	0.36	.666
Control vs. intervention \times Pre-intervention Score	-0.04 [-0.75, 0.42]	0.30	-0.12	.919
Control vs. intervention \times SES	-9.4e-03 [-0.46, 0.61]	0.27	-0.03	.969
Control vs. intervention \times No. sessions	2.2e-03 [-0.05, 0.05]	0.02	0.09	.822

Note. The model did not provide a significant fit to the data, indicating that our predictors, as a whole, did not capture significant variance in the outcome measure: $F(11, 35) = 1.1 [-4.18, 2.51]$, $p = .388$, $R^2 = .2574$, $N = 47$. SES = socioeconomic status.

indicates that the training delivered at the pre-intervention session was effective at boosting interactive shared reading behaviors. This finding is in accordance with the work of Dowdall et al. (2019), who reported that shared book reading interventions yield changes with large effect sizes on caregiver book sharing behavior.

However, contrary to our prediction, this increase in interactive shared book reading behaviors in the dialogic and pause reading conditions did not have an impact on the child's expressive and receptive language skills, their comprehension of syntax, or their MLU (although note that these results refer only to this intervention of a particular dosage and duration; longer interventions may yield more substantial results [see below]). The children in the intervention conditions, whose caregivers were taught particular interactive reading techniques, did not show a significant improvement on any of the language measures when compared to children in the control group, whose caregivers were simply instructed to read with their children. Power analyses and equivalence tests confirmed that these are likely to be true null effects for three of our language measures. The equivalence test result for the fourth measure (MLU) was

ambiguous, so we cannot determine whether meaningful differences (defined as a Cohen's d of .05 or above) exist between the groups in terms of their impact on MLU. However, our power analyses showed that these effects on MLU are so small that we would need substantially larger sample sizes to have enough power to detect significant effects at 80% power. Further research with a larger sample size is needed for MLU to draw robust conclusions, but even then, we expect effect sizes to be small.

Finally, also contrary to our prediction, there were no effects of SES. Children from both high- and low-SES backgrounds made equal gains in the language skills measured, and high- and low-SES caregivers implemented the interactive shared book reading interventions equally effectively. This is important as previous research has indicated that children from families of lower SES benefit less from shared reading interventions in terms of vocabulary and emergent literacy outcomes than their peers from families of higher SES (Manz et al., 2010; Mol et al., 2008). This discrepancy has led to a discussion in the literature about whether caregivers of lower SES are potentially less able to implement interactive

Table 10. Results of power simulations for the two main comparisons of interest for the four language outcome measures.

Outcome measure	Comparison	N	β level	N required for 80% power
Expressive language	Pause vs. dialogic	150	.180	942
	Control vs. pooled intervention		.190	4,272
Receptive language	Pause vs. dialogic	150	.122	1,056
	Control vs. pooled intervention		.033	9,078
Syntax comprehension	Pause vs. dialogic	150	.154	1,296
	Control vs. pooled intervention		.052	> 10,000
MLU	Pause vs. dialogic	48	.052	2,640
	Control vs. pooled intervention		.255	294

Note. N is the sample size in the current study. Power (β level) is the proportion of the simulations that yielded a p value of less than .05. N required for 80% power is the sample size required to observe a significant effect at these effect sizes 80% of the time. MLU = mean length of utterance.

Table 11. Results of the equivalence test for the two main comparisons of interest for the four language outcome measures.

Outcome measure	Comparison	<i>t</i>	<i>df</i>	<i>p</i>
Expressive language	Pause vs. dialogic	2.29	90.57	.012
	Control vs. pooled intervention	2.41	93.54	.009
Receptive language	Pause vs. dialogic	1.04	93.34	.152
	Control vs. pooled intervention	2.44	121.25	.008
Syntax comprehension	Pause vs. dialogic	1.84	73.81	.035
	Control vs. pooled intervention	-2.80	96.43	.003
MLU	Pause vs. dialogic	0.17	27.33	.432
	Control vs. pooled intervention	0.34	32.16	.369

Note. The inferential statistics here report the probability that we can reject the hypothesis that the effect size is at or above Cohen's $d = 0.5$. MLU = mean length of utterance.

shared book reading interventions than caregivers of higher SES. Contrary to this suggestion, our study found no evidence that low-SES caregivers are less able to implement interactive shared book reading interventions than caregivers of higher SES.

Turning to our language outcome measures, upon first inspection, the results appear to be at odds with the previous literature, which has indicated that shared book reading supports a wide range of early language skills, including vocabulary (e.g., Elley, 1989; Farrant & Zubrick, 2011), narrative and conversation skills (e.g., Morrow, 1988; Reese, 1995), print awareness (e.g., Justice & Ezell, 2000, 2004), future reading ability (e.g., Bus et al., 1995), and phonological awareness (e.g., Chow et al., 2008; Lefebvre et al., 2011). However, there are a number of possible explanations for the lack of significant effects on language in this particular study.

First, this study only tested two forms of interactive reading, so our results do not necessarily generalize to other strategies of interactive reading. It is possible that other forms of interactivity, not targeted in this study, may have an impact on early language development. However, note that our coding schemes capture exactly the kinds of reading behavior that are considered to be the most effective to boost language growth and, thus, that are promoted in most, if not all, interactive book reading training programs, both those for caregivers and those for early years teachers. Thus, we are confident that our results have some implications for interactive book reading advice more broadly.

Second, we did not test directly whether the caregivers changed their reading behavior during the intervention period, so we cannot conclusively state that the caregivers implemented the trained reading behavior. Instead, we have indirect evidence for this since we recorded how often caregivers read with their children during the intervention and whether they used the taught reading styles during the post-intervention reading session. The results indicate that the caregivers were able to adopt the reading style and that they maintained the knowledge of how to do the dialogic or pause reading through to the end of the 6-week intervention.

Third, the lack of an effect may be due to the use of an active control group rather than a passive or “business as usual” control group in this study. Passive control groups, in which the participants make no change to their behavior over the course of the intervention period, are more common in shared book reading interventions than active control groups (e.g., Chacko et al., 2018; Whitehurst et al., 1994). Active control groups require some change in the participants' behavior, which could be a change in an unrelated area (e.g., completing a play or craft activity) or a change in a related area (e.g., book reading). Active control groups provide a strong test of the effectiveness of a particular intervention technique as they allow the researcher to determine whether the specific content of the intervention is leading to improvement in the outcome measures. In this study, we asked caregivers in the control group to read according to a preset schedule (two books, 5 times a week) and to read a prescribed set of books. This amounted to a change in the caregivers' reading routine, which might itself have resulted in an improvement in the control group children's language.

A recent meta-analysis by Noble et al. (2019) found that the type of control group used in a study moderates the effect of shared book reading on language development. Noble et al. found that studies using a passive control group showed a small effect of shared book reading on language development ($\hat{g} = 0.231$, $p < .001$) but that studies using an active control group showed a negligible effect ($\hat{g} = 0.038$, $p = .584$). In the current study, we cannot determine whether the gains made by our active control group children were due to taking part in regular shared reading sessions or were due to simply taking part in an intervention (e.g., Hawthorne effect: change in participant behavior due to their awareness of being observed). However, in Noble et al.'s meta-analysis, they were able to comment on this, since the active control groups included groups exposed to nonlanguage-oriented interventions, such as structured play sessions or visuomotor skills training. The authors were thus able to speculate that shared book reading interventions in their current form may offer no more than a Hawthorne effect, although they were careful to be clear that this is not necessarily because shared book reading interventions

cannot support language development. Instead, they make a number of recommendations for the design of future research. These recommendations include carefully considering the outcome measures and the intervention dosage that would be required to lead to changes in these outcome measures.

This leads to the fourth possible explanation for why there is no difference in the language gains made by children in intervention and control groups in the current study, which is the combination of the outcome measures and the duration of the intervention. The chosen duration of 6 weeks was similar to that of previous studies that have reported positive findings of interactive reading on the language development of 2- to 3-year-old children (e.g., Chacko et al., 2018; Jacobi-Vessels, 2008; Lonigan et al., 1999; Valdez-Menchaca & Whitehurst, 1992; Whitehurst et al., 1994). Nevertheless, it is possible that the intervention needed a much higher dose to lead to changes in our outcome measures (see Dowdall et al., 2019, for a similar suggestion). Other types of intervention, albeit often conducted with older children, sometimes cover a long period of early childhood and last for a number of years (e.g., Barnes & Puccioni, 2017; DeBaryshe, 1993; Farrant & Zubrick, 2013; Shahaeian et al., 2018). Future work should investigate the effect of longer, more intensive dialogic and pause reading interventions.

The fifth, and final, explanation for why the interventions did not work may simply be that teaching caregivers to read interactively is no more effective than simply asking caregivers to read more with their children. In other words, it may be enough simply to increase the amount of shared book reading, as we did with our active control group, without having to teach caregivers to read interactively. However, we caution against dismissing all interactive shared book reading techniques on the basis of one study alone. It would be premature to come to this conclusion, given that interactive shared book reading programs often include a range of potentially language-boosting behaviors that have been linked to language development in previous research. For example, the child-directed speech delivered during interactive shared book reading contains higher levels of syntactic and lexical diversity than the speech children are exposed to during play-based activities (Cameron-Faulkner & Noble, 2013; Noble et al., 2018), and we know that high levels of syntactic and lexical diversity in speech directed to children are linked to higher levels of syntactic and lexical diversity in children's speech (Huttenlocher et al., 2002). Interactive shared book reading has also been shown to foster higher levels of joint attention, responsiveness, and contingent talk, all of which have been shown to support language development (Carpenter et al., 1998; Farrant & Zubrick, 2013; McGillion et al., 2017; Tomasello & Farrar, 1986). It also encourages the caregiver to use additional behaviors, which have all been shown to support children's language development, including expanding, recasting, and asking open-ended questions (Baker & Nelson, 1984; Cleave et al., 2015; Farrar, 1990; Girolametto & Weitzman, 2002; Huttenlocher et al., 2002, 2010; Nelson,

1977). Given these findings, it would be surprising if interactive shared book reading interventions that contain these language-boosting elements had no positive impact on children's language outcomes, although it is certainly possible it has an impact on some but not all language outcomes (e.g., vocabulary but not alphabets; U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2007, 2015).

Limitations

Although this study used a gold-standard clinical trial design and was preregistered on clinicaltrials.gov (NCT02625584), there were a number of limitations that should be addressed in future studies. Most importantly, we only tested two forms of interactive reading and only in comparison with an active control group of children whose caregivers were also instructed to read with their children, although they were given no training in interactive reading. Thus, it is possible that other forms of interactivity, which were not targeted in this study, may have an impact on early language development. It is also possible that the two intervention programs had an impact on the language outcomes but did not have a stronger effect than the effect in an active control group of children whose caregivers were also instructed to read regularly. Further work must explore these possibilities.

In addition, future work should investigate the effect of parental reading behaviors in more detail. We did not test directly whether the caregivers changed their reading behavior during the intervention period, so we cannot conclusively state that the caregivers implemented the trained reading behavior. Nor can we explore how caregivers implemented the trained reading behaviors, since we simply recorded the presence or absence of trained behaviors, not what those behaviors were; which behaviors were most common; and which were produced by caregivers in both intervention groups (e.g., prompts). More detailed analyses are required to gain a better understanding of how the training influenced caregiver reading behavior.

Finally, further work should investigate the effect of intervention duration and outcome measure choice. With regard to duration, it will be important to study the effect of longer, more intensive dialogic and pause reading interventions on the language development of 2- to 3-year-old children. With regard to outcome measures, future work could use different expressive and receptive language measures to determine whether more sensitive measures might pick up on more subtle changes (e.g., measures that target the words modeled in the books to determine whether shared book reading has more specific effects on vocabulary learning). Similarly, different measures of syntactic ability should be tested. Like all analyses of naturalistic speech data, our MLU measure relied on transcribers accurately representing child speech, which can be difficult for children of this age. In addition, since transcribing speech is time consuming, we were not able to include all children in the

MLU analysis. Thus, it is important to replicate the study using different measures of syntactic productivity.

Conclusion

This randomized controlled trial showed that caregivers from all socioeconomic backgrounds successfully adopted two types of interactive shared reading techniques. However, while both techniques were effective at increasing caregivers' use of interactive shared book reading behaviors, neither had an impact on the children's language skills over and above the language gains made by children in an active reading control group. Children in the intervention conditions did not show a significantly greater improvement on any of the language measures than the children in the control group. Note, however, that this study evaluated only two types of interactive shared reading programs, over a limited duration (6 weeks). Thus, we caution against coming to a more general conclusion that interactive shared book reading itself does not support language development. Instead, we make a series of recommendations for researchers and clinical professionals who are involved in designing and implementing such interventions for caregivers:

1. **Active and passive controls.** When evaluating the effect of a particular intervention, the use of active control groups controls for confounds such as Hawthorne effects. The specific content of the active control group will be important and will depend on the aim of the research. Non-book reading active control groups can help determine the effect of shared reading in general. Specific types of shared book reading can be compared, as in the study here, to determine whether some are superior to others for particular outcomes.
2. **Intervention duration.** Interventions should be of a duration and dosage likely to yield an effect size big enough to justify the cost of the intervention and should be calibrated to the outcome measures of the study. High-dose interventions are likely to be needed to have a measurable impact on some language skills.
3. **Outcome measures.** We should continue to investigate a range of language outcomes to find the outcomes that are best supported by shared reading. Clinical practitioners are encouraged to incorporate robust tests of an intervention's effectiveness using different outcome measures into their practice, to help inform the research literature. This recommendation is based on previous meta-analyses that have indicated that the efficacy of shared book reading interventions may vary depending on the specific outcome variables measured (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse, 2007, 2015).

Acknowledgments

This research was funded by Economic and Social Research Council (ESRC) Grant ES/M003752/1, awarded to Caroline Rowland and Thea Cameron-Faulkner. Caroline Rowland and

Thea Cameron-Faulkner are members of the ESRC International Centre for Language and Communicative Development at the University of Liverpool, for which support from the ESRC (Grant ES/L008955/1) is gratefully acknowledged.

References

- Alcock, K., Meints, K., Rowland, C. F., Brelsford, V., Christopher, A., & Just, J. (2020). *The UK Communicative Development Inventories: Words and Gestures*. J&R Press.
- Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low and middle-income families. *Applied Psycholinguistics, 19*(2), 209–223. <https://doi.org/10.1017/S0142716400010043>
- Baker, N. D., & Nelson, K. E. (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. *First Language, 5*(13), 3–21. <https://doi.org/10.1177%2F014272378400501301>
- Barnes, E., & Puccioni, J. (2017). Shared book reading and preschool children's academic achievement: Evidence from the Early Childhood Longitudinal Study—Birth cohort. *Infant and Child Development, 26*(6), Article e2035. <https://doi.org/10.1002/icd.2035>
- Bishop, B. M. (1951). Mother-child interaction and the social behavior of children. *Psychological Monographs: General and Applied, 65*(11), i-34. <https://doi.org/10.1037/h0093633>
- Bornstein, M. H., Haynes, M. O., & Painter, K. M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language, 25*(2), 367–393. <https://doi.org/10.1017/S0305000998003456>
- Boyle, J., Gillham, B., & Smith, N. (1996). Screening for early language delay in the 18–36 month age-range: The predictive validity of tests of production, and implications for practice. *Child Language Teaching and Therapy, 12*(2), 113–127. <https://doi.org/10.1177/026565909601200202>
- Bus, A. G., Leseman, P. P. M., & Keultjes, P. (2000). Joint book reading across cultures: A comparison of Surinamese-Dutch, Turkish-Dutch, and Dutch parent-child dyads. *Journal of Literacy Research, 32*(1), 53–76. <https://doi.org/10.1080/10862960009548064>
- Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*(1), 1–21. <https://doi.org/10.3102/00346543065001001>
- Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and child directed speech. *First Language, 33*(3), 268–279. <https://doi.org/10.1177/0142723713487613>
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i-174. <https://doi.org/10.2307/1166214>
- Chacko, A., Fabiano, G. A., Doctoroff, G. L., & Fortson, B. (2018). Engaging fathers in effective parenting for preschool children using shared book reading: A randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology, 47*(1), 79–93. <https://doi.org/10.1080/15374416.2016.1266648>
- Chow, B. W.-Y., McBride-Chang, C., Cheung, H., & Chow, C. S.-L. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology, 44*(1), 233–244. <https://doi.org/10.1037/0012-1649.44.1.233>

- Colmar, S. (2011). A book reading intervention with mothers of children with language difficulties. *Australasian Journal of Early Childhood*, 36(2), 104–112. <https://doi.org/10.1177/183693911103600214>
- Colmar, S. H. (2014). A parent-based book-reading intervention for disadvantaged children with language difficulties. *Child Language Teaching and Therapy*, 30(1), 79–90. <https://doi.org/10.1177/0265659013507296>
- Cleave, P. L., Becker, S. D., Curran, M. K., Owen Van Horne, A. J., & Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology*, 24(2), 237–255. https://doi.org/10.1044/2015_AJSLP-14-0105
- Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, 90(3), 759–773. <https://doi.org/10.1111/cdev.12974>
- DeBaryshe, B. D. (1993). Joint picture-book reading correlates of early oral language skill. *Journal of Child Language*, 20(2), 455–461. <https://doi.org/10.1017/S0305000900008370>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, 42(6), 1432–1443. <https://doi.org/10.1044/jslhr.4206.1432>
- Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2019). Shared picture book reading interventions for child language development: A systematic review and meta-analysis. *Child Development*. Advance online publication. <https://doi.org/10.1111/cdev.13225>
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24(2), 174–187. <https://doi.org/10.2307/747863>
- Farrant, B. M., & Zubrick, S. R. (2011). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language*, 32(3), 343–364. <https://doi.org/10.1177/0142723711422626>
- Farrant, B. M., & Zubrick, S. R. (2013). Parent-child book reading across early childhood and child vocabulary in the early school years: Findings from the Longitudinal Study of Australian Children. *First Language*, 33(3), 280–293. <https://doi.org/10.1177/0142723713487617>
- Farrar, M. J. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language*, 17(3), 607–624. <https://doi.org/10.1017/S0305000900010904>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), i-185. <https://doi.org/10.2307/1166093>
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools*, 33(4), 268–281. [https://doi.org/10.1044/0161-1461\(2002\)022](https://doi.org/10.1044/0161-1461(2002)022)
- Hammer, C. S., Nimmo, D., Cohen, R., Draheim, H. C., & Johnson, A. A. (2005). Book reading interactions between African American and Puerto Rican Head Start children and their mothers. *Journal of Early Childhood Literacy*, 5(3), 195–227. <https://doi.org/10.1177/1468798405058683>
- 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(1), 4–14. <https://doi.org/10.1037/a0027238>
- Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, 38(4), 271–278. <https://doi.org/10.1016/j.jcomdis.2005.02.003>
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3), 337–374. [https://doi.org/10.1016/S0010-0285\(02\)00500-5](https://doi.org/10.1016/S0010-0285(02)00500-5)
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, 61(4), 343–365. <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- Jacobi-Vessels, J. (2008). *The effects of a phonological awareness intervention on the decoding skills of kindergarten children* (Publication No. 3352038) [Doctoral dissertation, University of Louisville]. ProQuest Dissertations & Theses Global.
- Justice, L. M., & Ezell, H. K. (2000). Enhancing children's print and word awareness through home-based parent intervention. *American Journal of Speech-Language Pathology*, 9(3), 257–269. <https://doi.org/10.1044/1058-0360.0903.257>
- Justice, L. M., & Ezell, H. K. (2004). Print referencing: An emergent literacy enhancement strategy and its clinical applications. *Language, Speech, and Hearing Services in Schools*, 35(2), 185–193. [https://doi.org/10.1044/0161-1461\(2004\)018](https://doi.org/10.1044/0161-1461(2004)018)
- Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2020). Improving inferences about null effects with Bayes factors and equivalence tests. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 75(1), 45–57. <https://doi.org/10/gdk2z5>
- Law, J., McBean, K., & Rush, R. (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *International Journal of Language & Communication Disorders*, 46(6), 657–664. <https://doi.org/10.1111/j.1460-6984.2011.00036.x>
- Lefebvre, P., Trudeau, N., & Sutton, A. (2011). Enhancing vocabulary, print awareness and phonological awareness through shared storybook reading with low-income pre-schoolers. *Journal of Early Childhood Literacy*, 11(4), 453–479. <https://doi.org/10.1177/1468798411416581>
- Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, 108(1), 1–24. <https://doi.org/10.1016/j.jecp.2010.07.002>
- Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language & Communication Disorders*, 37(1), 3–15. <https://doi.org/10.1080/13682820110089911>
- Lonigan, C. J., Anthony, J. L., Bloomfield, B. G., Dyer, S. M., & Samwel, C. S. (1999). Effects of two shared-reading interventions on emergent literacy skills of at-risk pre-schoolers. *Journal of Early Intervention*, 22(4), 306–322. <https://doi.org/10.1177/105381519902200406>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Erlbaum.
- 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(4), 409–431. <https://doi.org/10.1016/j.ecresq.2010.03.002>
- Martin, P., & Bateson, P.** (1986). *Measuring behaviour: An introductory guide* (2nd ed.). Cambridge University Press.
- McCarthy, S. J.** (1997). Connecting home and school literacy practices in classrooms with diverse populations. *Journal of Literacy Research*, 29(2), 145–182. <https://doi.org/10.1080/10862969709547955>
- McGillion, M., Pine, J. M., Herbert, J. S., & Matthews, D.** (2017). A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. *The Journal of Child Psychology and Psychiatry*, 58(10), 1122–1131. <https://doi.org/10.1111/jcpp.12725>
- Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H.** (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, 19(1), 7–26. <https://doi.org/10.1080/10409280701838603>
- Mooney, J., Winter, K., & Connolly, P.** (2016). Effects of a book gifting programme on literacy outcomes for foster children: A randomised controlled trial evaluation of the Letterbox Club in Northern Ireland. *Children and Youth Services Review*, 65, 1–8. <https://doi.org/10.1016/j.childyouth.2016.03.009>
- Morrow, L. M.** (1988). Young children's responses to one-to-one story readings in school settings. *Reading Research Quarterly*, 23(1), 89–107. <https://doi.org/10.2307/747906>
- Nelson, K. E.** (1977). Facilitating children's syntax acquisition. *Developmental Psychology*, 13(2), 101–107. <https://doi.org/10.1037/0012-1649.13.2.101>
- Noble, C. H., Cameron-Faulkner, T., & Lieven, E.** (2018). Keeping it simple: The grammatical properties of shared book reading. *Journal of Child Language*, 45(3), 753–766. <https://doi.org/10.1017/S0305000917000447>
- Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J.** (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*. Advance online publication. <https://doi.org/10.1016/j.edurev.2019.100290>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A.** (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Office for National Statistics, National Records of Scotland, & Northern Ireland Statistics and Research Agency.** (2016). *2011 Census aggregate data*.
- Olson, W. C.** (1929). *The measurement of nervous habits in normal children*. University of Minnesota Press.
- Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K.** (2018). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*. Advance online publication. <https://doi.org/10.1016/j.ecresq.2018.04.001>
- Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E.** (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, 76(4), 763–782. <https://doi.org/10.1111/1467-8624.00498-i1>
- Purpura, D. J., Napoli, A. R., Wehrspann, E. A., & Gold, Z. S.** (2017). Causal connections between mathematical language and mathematical knowledge: A dialogic reading intervention. *Journal of Research on Educational Effectiveness*, 10(1), 116–137. <https://doi.org/10.1080/19345747.2016.1204639>
- R Core Team.** (2018). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.
- Reese, E.** (1995). Predicting children's literacy from mother–child conversations. *Cognitive Development*, 10(3), 381–405. [https://doi.org/10.1016/0885-2014\(95\)90003-9](https://doi.org/10.1016/0885-2014(95)90003-9)
- Richards, M. P. M., & Bernal, J. F.** (1972). An observational study of mother–infant interaction. In N. Blurton Jones (Ed.), *Ethological studies of child behaviour* (pp. 175–198). Cambridge University Press.
- RStudio Team.** (2019). *RStudio: Integrated development for R* [Computer software]. <http://www.rstudio.com/>
- Schulz, K. F., Altman, D. G., & Moher, D.** (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, 698–702. <https://doi.org/10.1136/bmj.c332>
- Shahacian, A., Wang, C., Tucker-Drob, E., Geiger, V., Bus, A. G., & Harrison, L. J.** (2018). Early shared reading, socioeconomic status, and children's cognitive and school competencies: Six years of longitudinal evidence. *Scientific Studies of Reading*, 22(6), 485–502. <https://doi.org/10.1080/10888438.2018.1482901>
- Tomasello, M., & Farrar, M. J.** (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463. <https://doi.org/10.2307/1130423>
- Tomasello, M., & Stahl, D.** (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, 31(1), 101–121. <https://doi.org/10.1017/S0305000903005944>
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse.** (2007). *Early childhood education intervention report: Dialogic reading*. <http://whatworks.ed.gov>
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse.** (2015). *Early childhood education intervention report: Shared book reading*. <http://whatworks.ed.gov>
- Valdez-Menchaca, M. C., & Whitehurst, G. J.** (1992). Accelerating language development through picture book reading: A systematic extension to Mexican day care. *Developmental Psychology*, 28(6), 1106–1114. <https://doi.org/10.1037/0012-1649.28.6.1106>
- Waldfoegel, J., & Washbrook, E. V.** (2010). *Low income and early cognitive development in the U.K.: A report for the Sutton Trust*. Sutton Trust.
- Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E.** (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, 30(5), 679–689. <https://doi.org/10.1037/0012-1649.30.5.679>
- Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A. C., Crone, D. A., & Fischel, J. E.** (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology*, 86(4), 542–555. <https://doi.org/10.1037/0022-0663.86.4.542>
- Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M.** (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24(4), 552–559. <https://doi.org/10.1037/0012-1649.24.4.552>
- Wiig, E. H., Secord, W., & Semel, E. M.** (2006). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2 UK)*. Harcourt Assessment.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E.** (2014). *Preschool Language Scale–Fifth Edition (PLS-5 UK)*. Pearson. <https://doi.org/10.1037/t15141-000>