Journal of Child Language (2020), **47**, 655–679
doi:10.1017/S0305000919000801

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Can inferencing be trained in preschoolers using shared book-reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability

Catherine DAVIES[1]*, Michelle MCGILLION[2,3], Caroline ROWLAND[4,5], and Danielle MATTHEWS[2]

[1]University of Leeds, UK, [2]University of Sheffield, UK, [3]Universityof Warwick, UK, [4]ESRC LuCiD Centre & Department of Psychological Sciences, University of Liverpool, UK, and [5]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
*Corresponding author: School of Languages, Cultures and Societies, University of Leeds, LS2 9JT. E-mail: c.n.davies@leeds.ac.uk

(Received 8 February 2019; revised 22 July 2019; accepted 1 October 2019; first published online 20 December 2019)

Abstract
The ability to make inferences is essential for effective language comprehension. While inferencing training benefits reading comprehension in school-aged children (see Elleman, 2017, for a review), we do not yet know whether it is beneficial to support the development of these skills prior to school entry. In a pre-registered randomised controlled trial, we evaluated the efficacy of a parent-delivered intervention intended to promote four-year-olds' oral inferencing skills during shared book-reading. One hundred children from socioeconomically diverse backgrounds were randomly assigned to inferencing training or an active control condition of daily maths activities. The training was found to have no effect on inferencing. However, inferencing measures were highly correlated with children's baseline language ability. This suggests that a more effective approach to scaffolding inferencing in the preschool years might be to focus on promoting vocabulary to develop richer and stronger semantic networks.

Keywords: randomised controlled trial; inferencing; language intervention; oral language; shared book-reading

The importance of inferencing skills

To make sense of language, children must make inferences. For example, they may need to infer what a pronoun (e.g., *it*) refers to, or why a protagonist in a story acted in a certain way, based on information distributed through the discourse. Inferencing skills are crucial for language comprehension because speakers and writers leave much of the content of their messages implicit. For example, on hearing that a character entered a room and turned on her torch, good comprehenders readily infer that the room was dark. In this case the missing information is provided by general world knowledge, which is integrated with information from the discourse as it unfolds. In this study, we evaluate the efficacy of a novel language intervention intended to promote four-year-olds' oral inferencing skills. This parent-delivered intervention was designed to prompt children's inferential thinking by giving them practice in answering inferencing questions during shared book-reading.

Inferencing skills allow comprehenders to construct a full and accurate representation of texts by linking events and working out causes and consequences of actions to create a coherent mental representation. Without good inferencing skills that draw from knowledge removed from the here-and-now to fill in implicit information, we cannot make sense of extended discourses such as narratives or instructions. When children start school, they face a sharp increase in the amount and range of the decontextualized language they hear (Hindman, Connor, Jewkes, & Morrison, 2008; Rowe, 2013), meaning that inferencing skills are in greater demand. Given the importance of good oral language at school, improving inferential language during the preschool years is likely to benefit school readiness when children start formal primary education. Strong inferencing skills across oral and written modalities can help enable a child to fully access the curriculum. More broadly, current educational policy emphasises the need for greater language comprehension skills (Law *et al.*, 2017; Oxford University Press, 2018), so inferencing remains a priority in primary education.

As children progress through the primary school years, inferencing becomes particularly important for reading comprehension and for related academic success. Indeed, much of the literature on inference-making comes from studies of reading comprehension (e.g., Cain & Oakhill, 1999; McGee & Johnson, 2003; Silva & Cain, 2015; Yuill & Oakhill, 1988; see Elleman, 2017, for a meta-analytic review). Several studies have found that children with poor reading comprehension are less likely to make inferences when reading than those with good comprehension (Cain & Oakhill, 1999; Oakhill, 1984), and a range of approaches including classroom-intervention and individual differences methodologies have reinforced the link between inferencing and reading comprehension. Text-based inference training has been effective in enhancing comprehenders' reading abilities (Bos, De Koning, Wassenburg, & van der Schoot, 2016; McGee & Johnson, 2003; Yuill & Oakhill, 1988; though note that these studies are not randomised controlled trials), and latent inferencing skill has been found to predict reading comprehension (Cain & Oakhill, 1999; Oakhill & Cain, 2012; Silva & Cain, 2015). More broadly, higher-level comprehension processes including inferencing account for unique variance in reading comprehension (Language and Reading Research Consortium [LARRC], 2017). Thus, there is a good evidence base showing that good text-based inferencing abilities provide a firm basis for later reading success.

Training inferencing skills

Although several studies have shown that it is possible to train inferencing in school-aged children to improve reading comprehension (Bos *et al.*, 2016; Clarke, Snowling, Truelove, & Hulme, 2010; McGee & Johnson, 2003, Yuill & Oakhill, 1988), very little is known about whether and how inferencing can be supported earlier on. Given the strong link between inferencing and reading skills, oral inferencing in the preschool years should help with the later demands of formal literacy education. Inferencing practice may also provide early protection to children at risk of becoming poor comprehenders, since a proportion of children may start school at risk of reading difficulties "not because they have problems with decoding or literal comprehension (although they may have these difficulties, too), but because they have not had extensive exposure to text inferencing that supports later, higher levels of literacy" (van Kleeck, 2006, p. 279). Although there have been many arguments in favour of promoting language skills in the preschool years, and successful interventions for doing so (e.g., Burgoyne, Gardner, Whiteley, Snowling, & Hulme, 2018), and although inferencing skills for oral language have been monitored in the preschool years (e.g., Das Gupta & Bryant, 1989; Filiatrault-Veilleux, Bouchard, Trudeau, & Desmarais, 2016; Pyykkönen, Matthews, & Järvikivi, 2010; Schulze, Grassmann, & Tomasello, 2013), how those skills can be strengthened BEFORE CHILDREN LEARN TO READ is currently unknown.

In a cross-sectional study with four- to six-year-olds, Florit, Roch, and Levorato (2011) found that inferencing skills play a specific role in oral language comprehension. In preschoolers, only three studies to our knowledge have explored whether it is possible to train inferencing skills – one educational and two clinical. First, in a 3-year, quasi-experimental (i.e., non-randomised) study beginning when a large sample of children were almost four years of age, Bianco *et al.* (2010) found improved oral comprehension as a result of regular, long-term, explicit, well-defined, comprehension-focused activities including inferencing. However, this study in preschools had a broader focus on comprehension skills more generally and thus we do not yet know which activities specifically supported inference-making. Second, in an 8-week oral inferencing training programme with preschoolers with language impairment, van Kleeck, Vander Woude, and Hammett (2006) reported that their training group outperformed non-intervention controls on receptive and inferential language (though with a small sample size of 15 children in each of the two groups). More recently, in a small-group book-sharing intervention with Australian pre-primary five- to six-year-olds with developmental language disorder, a randomised controlled trial found that children who had undergone oral inferential comprehension training ($n = 19$) showed an increase in inferential comprehension scores immediately after the 8-week intervention, maintained 8 weeks later. This group also scored higher than the control group for inferential comprehension on a post-intervention assessment of their ability to generalise inferential skills to new narrative contexts (Dawes, Leitão, Claessen, & Kane, 2019).

Although these three studies provide tentative evidence that building inferencing skills can improve oral language comprehension, and that inferencing ability can be trained under certain conditions during the preschool years, it is difficult to draw conclusions due to the diverse nature of the populations and the methodologies used. Until now, no study has investigated whether focused practice in inferencing, delivered as part of typically developing preschoolers' regular activities at home, will lead to improved inferencing skill in a large and diverse sample. Evidence from such a study would have clear implications for the way that inferencing is supported in the preschool years.

The studies that have found improvements in inferencing ability in school-aged children have used a wide range of instruction methods, from explicit teaching (Bos *et al.*, 2016; Clarke *et al.*, 2010; McGee & Johnson, 2003) to more implicit practice, e.g., asking comprehension questions about texts and allowing children to naturally discuss their answers with their peers (Yuill & Oakhill, 1988). In line with the literature, we define explicit vs. implicit instructional methods, respectively, as (i) guided activities that focus a child's attention explicitly on the pieces of information required for making an inference, and on the process of integrating them; and (ii) activities that elicit inferencing processes from the child incidentally through comprehension questions (Connor, Morrison, & Katch, 2004; Snow, 2001). As defined, the intervention reported in the current study uses implicit methods.

As we turn our focus to inferencing training in preschoolers, the range of suitable training methods narrows, since many forms of explicit instruction require an explicit understanding of the components of the skill being taught, e.g., the separability of discrete chunks of information, or of information sources. However, a wide range of implicit methods remains open for this age range: it has been suggested that younger children would particularly benefit from supportive dyadic contexts for inferencing, where they are encouraged to demonstrate their inferencing abilities via narratives rather than undergoing formal question-and-answer tests (van Kleeck, 2006, p. 292). Similarly, van Kleeck (2008) has suggested that one of the best ways to promote inferencing ability in younger children is to give them practice in making inferences by responding to questions about a story and then discussing answers. To our knowledge, these recommendations have yet to be taken up in a rigorous trial. Here we explore the value of parent–child book-reading as a basis for this kind of practice. Specifically, we test whether practising making inferences in order to respond to caregiver questions during shared book-reading promotes four-year-olds' inferencing ability.

Intervention approach

Strengthening inferencing skills in the early years is likely to have advantages for oral language comprehension and later reading ability. Despite evidence showing that inferences can be trained in school-aged children using a range of methods, from answering inferencing questions to formal explicit teaching, we do not yet know: (i) whether training inferencing in the preschool years is possible; and (ii) if so, whether it is possible in this age group via implicit methods. This is particularly important, since formal, explicit instruction methods rely on an understanding of the subcomponents of inferencing so may not be easily accessible for this age group. To address this gap, we report the results of a RANDOMISED CONTROLLED TRIAL (RCT) to test the effect of a parent-delivered intervention that gave four-year-old children practice in responding to inferencing questions during book sharing. To provide the best chance of success, our design combined elements of successful inferencing interventions for older children with recommendations for scaffolding inferencing in preschoolers. It follows Yuill and Oakhill's (1988) finding that comprehension questions improve inferencing ability, as well as van Kleeck's (2008) evidence-based recommendations for fostering inferential language in preschoolers, e.g., embedding scripted inferencing questions, prompts, and feedback in shared reading materials (Ard & Beverly, 2004; Karweit, 1989; van Kleeck *et al.*, 2006). Inserting these prompts ahead of time increases the amount of 'thinking aloud' between dyads (Kucan & Beck, 1997), and improves fidelity.

Our intervention was designed to prompt younger children in their inferential thinking. Although preschoolers are able to engage in inferencing, evidence suggests that they are less likely than their older peers to do so spontaneously (Florit *et al.*, 2011). Through naturalistic questioning (based on evidence showing that some parents naturally engage in literal and inferential talk during shared book-reading;

Hammett, van Kleeck, & Huberty, 2003), our training highlighted the fact that there is information to be had that is not explicitly stated, and encouraged children to fill in the gaps using clues provided in the text or from their prior knowledge. By raising their awareness of these gaps, children were alerted that an inference needed to be made, and encouraged to strive for coherence (Cain & Oakhill, 1999, p. 501). Further, unlike studies involving classroom-based, group training sessions that use explicit training methods to highlight textual cues to implicit meaning (Bianco *et al.*, 2010; McGee & Johnson, 2003; Yuill & Oakhill, 1988; Zucker, Justice, Piasta, & Kaderavek, 2010), the intervention was run at home by parents, meaning that, if successful, the programme could be adopted without the need for specialist training.

Shared book-reading was chosen as the medium for the intervention for several reasons. Children who read regularly with an adult in the preschool years learn language faster, enter school with a larger vocabulary, and become more successful readers in school (Mol, Bus, de Jong, & Smeets, 2008). Shared book-reading facilitates more complex talk than traditional caretaking or play activities (Snow, 1993), and exposes children to vocabulary and syntactic structures beyond what they would hear in everyday speech (Cameron-Faulkner & Noble, 2013; Hoff-Ginsberg, 1991). Thus, shared book-reading is a potentially powerful tool for supporting the development of vocabulary, narrative and conversational skills, complex syntax, and other literacy practices such as print and phonological awareness (Burgess, 2010; Ezell & Justice, 2000). More specifically, our intervention asked open-ended questions; a technique from dialogic reading interventions. Dialogic reading encourages caregivers to be more responsive to the child during shared book-reading, and in general has been shown to have a positive impact on a child's oral language development (Baker & Nelson, 1984; Cleave, Becker, Curran, Van Horne, & Fey, 2015; Farrar, 1990; Girolametto & Weitzman, 2002; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Valdez-Menchaca & Whitehurst, 1992; though see a meta-analysis by Noble *et al.* (2019), and a randomised controlled trial by Noble *et al.* (unpublished observations) for evidence of no effect or small effects of dialogic reading on children's language skills). Most pertinently for our intervention, shared book-reading is a good medium for linking social conversations (e.g., about personal events and real-world knowledge) and text inferencing skills for two reasons. First, because some caregivers naturally ask their children questions about the shared story that require them to make inferences about the text, they model the kinds of information that support text comprehension, and then support the child in answering via various types of scaffolding (van Kleeck, 2006, 2008). Second, oral inferencing practice is particularly suited to shared book-reading because it takes place within the same activity that it will later be applied in when reading, i.e., generating meaning from information presented in books (van Kleeck, 2006, p. 275). Thus, we use shared book-reading as an activity that will provide the natural apprenticeship for later independent inferencing.

Design and hypotheses

The aim of this RCT was to test whether training parents to ask their children inference-eliciting questions during shared book-reading (and supporting them to do so with in-text questions) is effective for promoting inferencing ability in four-year-olds. The primary outcome measure was children's ability to answer inferencing questions after completing the 4-week intervention (controlling for baseline ability). The inferencing training group was compared with an active control group of children who spent the intervention period working through a maths exercise book with their caregiver. We hypothesised that the training group would make significantly greater gains in inferencing ability than the control group. The secondary outcome measure was the change in children's NFER Baseline Reception Assessment Language and Communication scale (National Foundation for Educational Research, 2015) (NFERL); a standardised assessment frequently used in British primary schools to gauge children's language ability upon school entry (aged four to five years). We did not have a hypothesis regarding potential effects on the NFERL scale as transfer is not often seen in response to cognitive training programmes (Sala & Gobet, 2017), but we were interested to assess this all the same.

Method[1]

This educational intervention was preregistered at <https://clinicaltrials.gov/ct2/home> (NCT02854462, 'Appendix A'). Ethical approval was granted by the Psychology Ethics sub-committee at the University of Sheffield.

Participants

The Consolidated Standards of Reporting Trials (CONSORT) diagram is reported in Figure 1 and the checklist appears as 'Appendix B' (Schulz, Altman, & Moher, 2010). One hundred four-year-olds (53 female) were recruited in the north of England from a volunteer database at the University of Sheffield's Department of Psychology (Mean age at Baseline = 50.3 months; Median = 50 months: Range: 48 to 56 months; Mean age at Post-test = 51.5 months; Median = 51 months: Range: 49 to 58 months). Eighty-three caregivers and their children had previously taken part in a separate randomised controlled trial investigating the role of caregiver contingent talk on early language development (McGillion, Pine, Herbert, & Matthews, 2017). These children did not differ on any measures collected at baseline from those who had not been involved in the previous study (n = 17). Participants were specifically recruited to be representative of the UK population in terms of SES: forty-five percent of households were not educated to degree level. Eighty-nine caregivers gave permission for their data to be uploaded to the UK Data Archive (UK Data Service. 10.5255/UKDA-SN-853233).

Inclusion criteria

Children were first-born (to control for potential birth-order effects), full-term (i.e., born no more than 3 weeks prematurely), with birth weight over 2.5 kg, and were monolingual English speakers (to allow for the administration of standardised language assessments). Exclusion criterion: neither caregivers nor children had any significant known physical, mental, or learning disability.

At baseline visits, families were given a cuddly toy and the materials required to complete the intervention, and a second cuddly toy and a £40 gift voucher on completing the post-test visit.

[1]An extended version of the methods section adhering to CONSORT guidelines can be accessed at <https://osf.io/95qr8>, along with all Appendices.

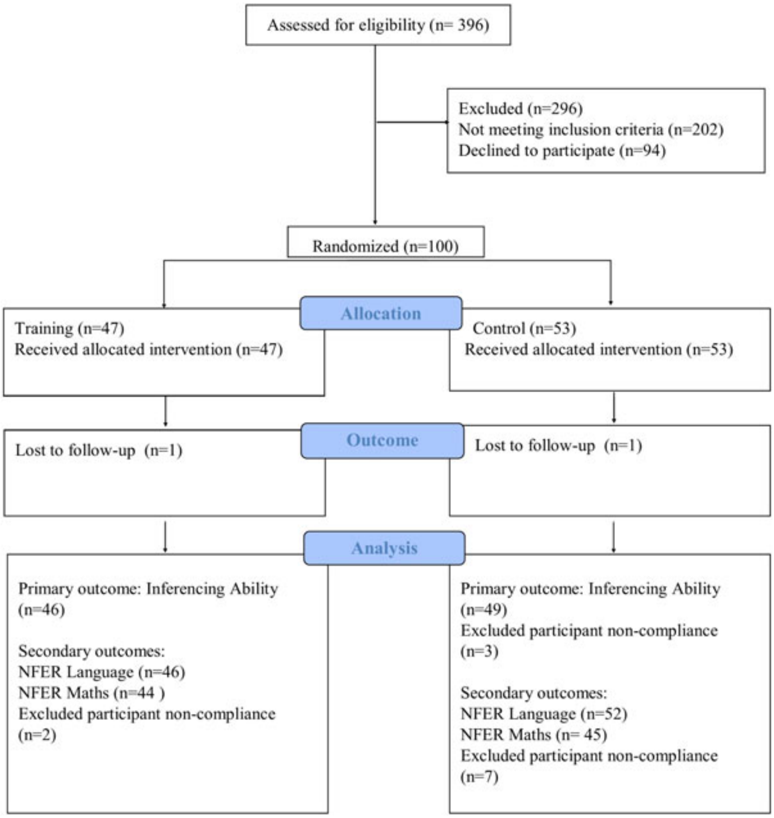

Figure 1. CONSORT flow diagram.

Materials

Intervention videos

A short video was used to deliver the Inferencing Training Intervention to caregivers. The script was developed by the authors to explain in lay terms what inferencing is and why it might be important for language and reading comprehension and, by extension, success in school ('Appendix C'). Stills and video clips, collected during piloting for this study, were used to illustrate how caregivers and their children might engage in inference-eliciting dialogues while reading books. This method of intervention administration has been used successfully in previous studies of language development (McGillion *et al.*, 2017), and was chosen for its consistency. Qualitative feedback in the exit questionnaire suggested that caregivers had enjoyed the reading comprehension video and found it useful in explaining the theory behind the study.

A second video (matched in length, format, production, and aims to the training condition) introduced caregivers to the Mathematical Control Intervention.

Intervention support materials

Children in the inferencing training condition were given 10 books. Inference-making questions were pasted alongside the text of these books to elicit inferencing during shared book-reading. Each question label included a picture of a tiger, who was introduced on the front cover of every book. Caregivers explained to the children that the tiger might need some help to understand the story, and that they could do this by answering the questions beside his picture throughout the story. Caregivers were encouraged to provide supportive individual feedback for correct responses (see information leaflet in ‘Appendix D’). For questions where the child did not respond, or responded incorrectly, model answers and feedback were included on the question labels, e.g.:

- Why does Percy need an extra blanket tonight?
- Perhaps he is trying to get warm.
- Can you remember what the weather was like outside? It was very cold!

One Snowy Night (Butterworth, 2011, p. 4).

See ‘Appendix E’ for book titles, inferencing questions, and model responses.

Video analysis during piloting confirmed that caregivers understood these instructions and were able to incorporate the question prompts and feedback into their usual book-reading routine (this was also endorsed in their oral feedback to us). Children in the control condition were given the commercially available maths workbook *At Home with Counting* (Ackland, 2012). This book introduces simple number knowledge (e.g., learning the numbers 1–10) and skills (e.g., sequencing, adding, more/less than) through matching drawing and colouring activities. Each page contained instructions for caregivers to encourage their child’s participation and support learning. Families in both conditions were given an intervention diary to record each time they read a particular book (inferencing training condition) or completed a page in the maths workbook (maths control condition) and to comment on their experience of taking part in these activities. Qualitative analysis of these comments after the intervention was complete suggested that caregivers understood what was expected of them during the intervention period.

Measures of inference-making

Age-appropriate story vignettes and questions were used to measure children’s inference-making ability at baseline and post-test. Inferencing vignettes for preschoolers taken from the Language and Reading Research Consortium (LARRC; see Currie & Cain, 2015; Language and Reading Research Consortium, 2015, for details of their construction and validation) were administered at baseline (Birthday) and at post-test (A New Pet; A Family Day Out Part 1). Additional author-designed vignettes followed the LARRC template (baseline n = 1, Rover the Dog (see Table 1); post-test n = 1; Jessie’s Birthday Party, ‘Appendix F’), and were designed to portray familiar scenarios that tapped into four-year-olds’ world knowledge.

To demonstrate comparability between the author-designed vignettes and those from the LARRC (2015) materials, the number of utterances, number of morphemes, word tokens (i.e., the total number of words including repetitions of the same word), and word types (i.e., number of different words) were computed using CLAN (Computerized Language Analysis; MacWhinney, 2000). Two measures of linguistic richness were used: global syntactic complexity (indexed by the mean length of utterances in morphemes) and lexical diversity (type:token ratio). These analyses suggest that the LARRC and author-designed vignettes were of a comparable level of difficulty.

Table 1. Baseline vignette *Rover the Dog* with inferencing questions

This story is called *Rover the Dog.* Listen carefully, and try to remember the story so that you can answer the questions.
Child is shown pictures of characters.
This is Rover, this is Jack, and this is Jack's Dad. So that's Rover, Jack, and Jack's Dad.
Jack and his Dad woke up early one Saturday morning. They went downstairs. Jack wanted a banana and an apple. Dad told him to look in the cupboard. He found the fruit, and then decided to go out with Dad and his dog Rover. They put on their wellies and opened the door. Dad said, "it's a good job we have our umbrellas isn't it!" He gave Rover a dog biscuit for being good and off they went.
1. Why did Jack and his Dad go downstairs? 2. Where were the banana and apple? 3. What was the weather like? 4. Who gave Rover a dog biscuit?
Let's see what happens next.
Jack's dog, Rover, loved playing in puddles. When they arrived, Jack played on the swings and the slide. Next, he went on the roundabout. All of a sudden, Rover ran off! Dad shouted at him to come back and Jack ran after him, worried. He thought Rover would get lost. Finally, Jack caught up with Rover and took him back to Dad. Dad said, "urgh he's all wet and muddy!". Jack smiled.
5. Where did they walk to with Rover? 6. Who thought Rover would get lost? 7. Why was Rover wet and muddy? 8. How did Rover feel when he was playing in the puddles?

Each short story was read aloud by the experimenter and was followed by between four and eight questions to assess inferencing ability. Questions followed the order in which the information was presented in each story vignette and required the child to integrate information across the text and/or their world knowledge to, for example, infer character motivations (e.g., Table 1, questions 1 and 5), emotions (question 8), and semantic (question 2) and anaphoric relationships (questions 4 and 6). We tested a range of inference types so that our results could inform interventions that would comprehensively promote the range of inferences that children face during oral comprehension. In this respect, our materials are in line with standardised measures of reading comprehension that frame inferencing as a broad construct, e.g., the Neale Analysis of Reading Ability (NARA II; Neale, 1989) and the York Assessment for Reading Comprehension (YARC; Hulme *et al*., 2009). Inferencing questions by type, with by-group scores, are in 'Appendix G'. Author-designed vignettes and questions were administered first at both time-points. These stories were presented in two parts to minimise memory demands, and included pictorial supports to illustrate characters in the story.

Procedure

Families meeting the eligibility criteria were invited to take part in a study investigating factors that impact on school readiness. Prior to this appointment, participating caregivers completed a Family Questionnaire to measure demographic information, e.g., caregiver education and household income (see Alcock, Meints, & Rowland, 2017, for details of its construction), and a Home Life questionnaire to collect information about literacy-related behaviours and attitudes e.g., how often someone read with the child in a typical week.

Randomisation

Dyads were randomised to either the inferencing training or maths control condition according to CONSORT 2010 guidelines (Schulz *et al.*, 2010). Randomisation was conducted by an independent statistician at the University of Sheffield. Randomisation was stratified by household education (degree or no degree) and the condition to which dyads had been allocated in a previous intervention study if they had taken part in it (McGillion *et al.*, 2017). If the family had not taken part in the prior intervention (n = 17), they were allocated a condition envelope for a family that had taken part in the prior intervention but who had declined to take part in this study, matching for SES. For each participant number, condition allocations were placed in a sealed envelope, identified only by participant number, by a research assistant not involved in any other aspect of the project. Another researcher who administered the baseline measures and the intervention became aware of condition allocation by opening the envelope during the baseline visit, and only once the final baseline measure had been collected. This ensured that baseline measures were collected blind to condition allocation. This research assistant, having opened the envelope with the appropriate participant number to find out which condition the dyad had been randomised to, administered the relevant intervention. Intervention groups did not differ as a function of child age, gender, or SES.

Baseline data collection

Caregivers and their children completed two baseline visits. On the first visit, at the university, children completed several measures of mathematical ability as part of a separate study on mathematical development (Yanez Diaz Barriga, 2018). The second visit took place in the family home. After two cameras had been set up (Sony HDR-PJ810E and Sony HDR-PJ220E) and turned on, caregivers and their child spent approximately 10 minutes completing a book-reading session as a warm-up activity before baseline data collection began. First, this involved collecting a measure of child inferencing ability. The researcher read two story vignettes (one author-designed; Rover the Dog (Table 1), the second from the LARRC; Birthday). Children were asked to listen carefully to the stories so that they could answer the inference-eliciting questions that followed each story. Aside from general encouragement, no other feedback was given. Then, child language and communication was measured by the researcher (secondary outcome) using the NFER Baseline Reception Assessment Language and Communication scale (NFERL) (National Foundation for Educational Research, 2015) and the Language Content index of the Clinical Evaluation of Language Fundamentals Preschool 2 UK (CELF) (Wiig, Second, & Semel, 2006). The NFERL assesses phonics, picture sequencing, story prediction, word reading, simple sentence reading, and name writing. The Language Content index of the CELF is a measure of vocabulary breadth, concept development, comprehension of simple and complex sentences, and comprehension of associations and relationships among words.

The intervention

After all baseline measures were collected, the researcher opened the envelope containing the dyad's condition allocation and administered the appropriate intervention.

Inference Training Condition

The researcher explained that the study was investigating whether asking questions during shared book-reading could help language comprehension before children start school. Caregivers were shown the intervention materials, watched the intervention video, and were asked to read each of the 10 books (with inferencing questions included) at least twice over the course of the following month (i.e., a minimum of 20 sessions), and were given a leaflet summarising the main intervention message ('Appendix D').

Mathematical Control Condition

The researcher explained that the study was investigating whether completing daily maths activities could help children get ready for school. Caregivers watched a video explaining what the intervention involved, were shown the maths workbook, and were asked to complete one or two pages a day over the course of the following month (i.e., a maximum of 20 sessions).

Caregivers in both conditions were given an intervention diary to record how often they completed the relevant intervention activities and their impressions of having done so.

Post-test data collection

Approximately one month later, caregivers and children visited the university for post-test data collection. A version of the Home Life questionnaire (adapted to include questions about the activities completed over the past month), and an exit questionnaire about the general experience of taking part in the study were posted to caregivers in advance of this visit. Caregivers were asked to complete these questionnaires and to bring it with their completed intervention diary to the university in a sealed envelope.

The researcher read three different short vignettes following the protocol established at baseline (see 'Materials' section: one author-designed; Jessie's Birthday Party, and two from the LARRC; A Family Day Out Part 1; A New Pet), see 'Appendix F'). These vignettes were of equivalent total length to those administered at baseline and were matched for story theme. After each story, the researcher asked a series of questions designed to measure the child's INFERENCE-MAKING ABILITY (primary outcome). The Communication Language and Literacy and Mathematical Literacy components of the NFER Baseline Reception Assessment were administered to measure CHILD LANGUAGE (secondary outcome) and mathematical ability.

Debrief

In accordance with ethical guidelines laid down by the University of Sheffield ethics committee, all caregivers were fully debriefed by email after all children had completed the final outcome visit.

Coding and measures

Inferencing ability at baseline and post-test was measured by child responses to the inference eliciting questions following story vignettes. Responses to each inference question were scored from video-recordings by a researcher blind to condition allocation. Correct responses that demonstrated full inference-making were awarded 2 points. Partially correct answers that lacked full inferencing scored 1 point.

Unintelligible responses, "I don't know", or incorrect responses scored 0. For example, for question 5 in Table 1; "Where did they walk to with Rover?", "The park / the playground" scored 2 points, "swings and slides / roundabout" scored 1, and "shopping / for a walk" scored 0. For any response that was scored as partially correct or incorrect, the researcher asked a background question(s) or prompted the child in line with the rubric to help the child follow the narrative, and so that subsequent questions could be administered. Complete scoring schemes are presented in 'Appendix F'. Scores for individual questions were summed to produce an overall inferencing score at baseline (out of a maximum score of 32 for the 16 questions at baseline) and post-test (out of a maximum score of 40 for the 20 questions at post-test). Rare instances of missing data were replaced with the sample mean for the particular item. At baseline, five participants had at least one missing datapoint on measures of inferencing ability, each with an average of 2.4 items missed out of the 16 items on this scale (totalling less than 1% of data on this measure). Four participants had at least one missing datapoint on post-test measures of inferencing ability, each with an average of 3 missed items out of the 20 items on this scale (totalling less than 1% of data on this measure). Incidences of missing data due to experimenter error or caregiver interference were replaced with the sample mean for the item in question.

Child language and communication ability was measured using the NFER Baseline Reception Assessment Language (baseline and post-test) and the CELF Language Content Index (baseline). These were scored from video-recordings by a researcher blind to condition allocation. A raw frequency score was calculated for each test according to individual assessment guidelines. Incidences of missing data due to experimenter error or caregiver interference were replaced with the sample mean for the item in question. Eight participants had at least one missing datapoint on the CELF, each with an average of 2.7 missed items out of the 59 items on this test (totalling less than 0.5% of data across the dataset on this measure). Nine participants had at least one missing datapoint on the baseline NFER, each with an average of 2.8 missed items out of the 43 items on this test (totalling 1% of data across the dataset on this measure). Twelve participants had at least one missing datapoint on the post-test NFER, each with an average of 1.8 missed items out of the 43 items on this test (totalling 0.5% of data across the dataset on this measure). Results for primary and secondary outcomes (as well as for the post-hoc analyses reported below) did not change when statistical models were run without the imputed data.

Reliabilities

Ten percent of responses to questions measuring inferencing ability, randomly selected, were double coded by a researcher blind to condition allocation at baseline ($n = 10$) and post-test ($n = 10$). Correlations between scorers indicated high levels of agreement at baseline ($r = 0.97$). There was 100% agreement at post-test. The CELF and NFER Language were coded live, using the standardised tests stopping rules. A second researcher blind to condition allocation recoded these tests from the video-recording to check the accuracy of the test administration and scoring. Internal consistency was acceptable for our main measure of inferencing ability at post-test ($\alpha = .76$). Baseline tests of inferencing ability had a Cronbach's alpha of .65. As a measure of the predictive validity of the inferencing measures, baseline and post-tests of inferencing were found to be positively correlated in the control group ($r = 0.59$).

Table 2. Means (SD) scores for Inference, CELF, and NFER language tests at baseline and post-test as a function of condition

		Inferencing training		Maths control	
	N	M	SD	M	SD
Baseline Inference raw /32	100	14.4	5.80	16.8	5.23
Baseline Inference %	100	44.9	18.1	52.6	16.4
Baseline CELF Language Content raw	98	47.6	9.10	50.2	8.21
Baseline NFER Language	100	17.6	6.82	18.3	7.20
Post-test Inference raw /40	95	20.1	7.21	20.9	6.71
Post-test inference %	95	50.2	18.0	52.3	16.8
Post-test NFER Language	98	20.1	7.56	19.9	6.68
Home sessions completed	79	23.9	7.92	13.1	8.03

Sample size and statistical methods

Sample size was calculated to detect a medium effect size for the primary outcome measure (inferencing ability) with 80% power at the 5% level of significance, allowing for up to 20% attrition rate. To compare primary and secondary outcomes across intervention groups, we fitted separate linear regression models to each outcome measure with condition and an equivalent baseline measure as predictors. All analyses were conducted using R version 3.3.1 (R Core Team, 2013) and RStudio Version 1.1.419 (RStudio Team, 2015). In two final post-hoc analyses we used correlation to explore individual differences between language (at baseline and post-test) and inferencing ability.

Results

Children were generally able to engage with the inferencing task, and scores on the baseline and post-test were normally distributed. Table 2 reports descriptive statistics for all baseline and post-test measures along with the number of sessions parents reported having completed at home for each condition. Percentage scores for the baseline and post-tests of inferencing were calculated by dividing total scores by the maximum possible score (32 at baseline and 40 at post-test), and are shown in Table 2 to facilitate interpretation: results of statistical analysis are reported for the raw scores only. Children in the maths control condition scored slightly but significantly higher on the baseline test of inferencing than children in the inferencing training condition ($t(98) = 2.23$, $p = .03$). In line with our statistical analysis plan, these baseline scores are controlled for in the analysis of the effect of the intervention below (Table 3). There was no significant difference between groups on either measure of language or communication collected at baseline (NFER Language $t(98) = 0.48$, $p = .63$; CELF $t(93) = 1.48$, $p = .14$). Both groups scored within expected norms on the CELF Language Content Index. Children in the inferencing training condition made bigger pre- to post-test numerical gains on the inference tests (45% at baseline and 50% at post-test) than children in the maths control condition (53% at baseline and 52% at post-test). Parents in the inferencing

Table 3. Regression model fitting condition and baseline inferencing to post-test inferencing (n = 95)

	B	SE	T	*p*
Intercept	20.01	0.89	22.52	< .001
Condition	0.85	1.29	0.65	.51
Baseline Inferencing	3.48	0.65	5.38	< .001

Note. $R^2 = .24$.

Table 4. Regression model fitting condition and NFER language baseline to post-test NFER language (n = 98)

	B	SE	T	*p*
Intercept	19.54	0.54	35.9	< .001
Condition	0.72	0.79	0.9	.37
Baseline NFER Language	5.93	0.40	14.9	< .001

Note. $R^2 = .70$.

condition reported completing sessions at home on more days than parents in the maths condition. This simply reflects the fact that, once the maths book had been completed, parents were less likely to return to it again, whereas the story-books for the inferencing training were often shared multiple times.

Effect of the intervention on primary and secondary outcome measures

To test for an effect of the intervention on inferencing scores (our primary outcome measure), we built a linear regression model with inferencing at post-test as the outcome variable and intervention condition plus inferencing at baseline (scaled and grand mean centered) as predictors. The model is reported in Table 3. Controlling for baseline, there was no statistically significant effect of condition on inferencing outcomes (Hedge's g = 0.14). Hedges' g (calculated using the R package metanalytic: Xiao, Kasim, & Higgins, 2016) is a corrected measure of effect size for continuous variables in smaller samples. It is interpreted in the same way as Cohen's d, i.e., 0.2 is a small effect, 0.5 is medium, and 0.8 is a large effect.

To test for an effect of the intervention on NFER language scores (our secondary outcome measure), we built a linear regression model with NFERL at post-test as the outcome variable and intervention condition plus NFERL at baseline (scaled and grand mean centred) as predictors. The model is reported in Table 4. Controlling for baseline, there was no statistically significant effect of condition on language outcomes (Hedge's g = 0.18).

To further explore these null effects, we ran equivalence tests on our primary (inferencing ability) and secondary (NFER language scores) outcome measures (Lakens, Scheel, & Isager, 2018). Equivalence testing is a variant of hypothesis testing that examines whether the difference between groups is more or less extreme than the smallest effect size of interest, i.e., are groups significantly equivalent. We used the two one-sided test procedure (TOST from R package TOSTER: Lakens, McLatchie, Isager, Scheel, & Dienes, 2018), setting the minimal effect size of interest at 0.5 (Colmar, 2014; Noble *et al.*, unpublished observations). These analyses showed that, at post-test, there was no meaningful difference between participants in the control and intervention groups with respect to both inferencing ($t(91) = -1.847$, $p = .034$) and language ability ($t(91) = 2.373$, $p = .001$).

Individual differences in inferencing and language ability

As an exploratory post-hoc analysis, we tested whether individual differences in inferencing were associated with language ability as measured by the CELF Language Content Index. Recall that the CELF measures vocabulary, concept development, and comprehension of sentences and of relationships between words – all key components of inferencing. A positive correlation would indicate that the inferencing is related to general language skill. Considering all participants at baseline, inferencing ability was significantly associated with performance on the CELF Language Content Index ($r = 0.47$, $p < .01$). At post-test, considering only children in the maths control condition (for whom there could have been no effect, however small, of the intervention), here too this measure of inferencing was highly correlated with children's baseline CELF Language Content Index ($r = 0.63$). These correlations are about the same size as the correlation between baseline and post-test measures of inferencing themselves ($r = 0.59$, again considering children in the control condition only, although correlations are of similar size for the full sample; $r = 0.49$). This suggests that our measure of inferencing is related to more general language ability.

Parents' responses to the intervention

Parents' qualitative comments about the inferencing training were extracted from the intervention diaries. Here we summarise the major trends, together with more general comments about study participation from an exit questionnaire.

Thirty-eight diaries were returned from the 47 families taking part in the inferencing training. These indicated that a mean of 23.9 sessions were completed over the duration of the intervention (SD = 7.8; Range 11–49). Recall that families were instructed to read each book at least twice during the month, so the minimum expected number of sessions was 20. Thus, we take the level of uptake and engagement in the training as moderate to high. Most parents saw value in preparing their children for the transition to school and reported that the activities were enjoyable, though they reported some difficulty fitting the sessions around other daily activities. Although parents commented that their children enjoyed certain books more than others, there was no difference in how often each of the 10 books was read (the mean number of reading sessions per book ranged from 2.3–2.8). Regarding children's levels of concentration during the shared reading sessions, some parents reported that their children enjoyed answering the questions, whereas others were distracted from the story by looking for the tiger stickers. Time of day was also cited as a factor in levels of tiredness and concentration. The repetition of books and questions elicited both positive and negative comments. Many parents were keen to report that the repetition strengthened their children's confidence and understanding of what was happening in the stories and, at least for the books that their children enjoyed, they were enthusiastic about repeated reading. However, a few parents also reported that children were frustrated by being asked the same questions. The main implication for our intervention is that children may not have been engaging in inferencing on subsequent sessions and instead either refusing to answer the question or rote responding from memory.

Of the range of inference types in the training materials, parents reported that their children found some harder than others. For example, inferences about why characters were feeling a certain way were challenging for some children, as were predictions. Inferencing was also sometimes hindered by a lack of world knowledge. For example, some children needed an explanation of the meanings of *mustard* or *carsick* in order to attempt the relevant inferencing question.

An unanticipated advantage of the inferencing training was its ability to give parents a means of explicitly assessing how much their children understood from a shared story. Many appreciated the chance to learn about their child's abilities. Some had underestimated how much their children understood but had revised their assessment from their child's responses to the inferencing questions. Together with the observations of spontaneous shared reading during the pilot study, these comments also suggest that the intervention went beyond parents' usual practices when reading with their children. The fact that parents would not typically ask this number and type of questions during shared book-reading means that our intervention was qualitatively different to business-as-usual for the majority of families in our sample. Although some parents (particularly those with higher levels of literacy; Bus, Leseman, & Keultjes, 2000; Heath, 1983) may engage in a lot of extra-textual talk when reading some genres (particularly information books; Anderson, Anderson, Lynch, Shapiro, & Kim, 2012; Pellegrini, Perlmutter, Galda, & Brody, 1990; Potter & Haynes, 2000; Price, van Kleeck, & Huberty. 2009), many do not. The socioeconomic diversity of our sample and the use of story-books in our study means that substantial differences are likely between the reading style imposed by our intervention and what the majority of parents in our sample would normally do.

Discussion

This randomised controlled trial evaluated a language intervention intended to boost inferencing skills using implicit training, delivered by parents to their preschool children. The training was designed to prompt four-year-olds' inferential thinking by giving them practice in answering inferencing questions during shared book-reading. The training had no significant effect on either inferencing skills or on language and communication skills. Despite good theoretical justification, high levels of engagement by the participating families, and a rigorous RCT design, our intervention did not effect significant change. Based on methods used in previous inferencing interventions that successfully improved comprehension in school-aged children (Bianco *et al.*, 2010; McGee & Johnson, 2003; van Kleeck *et al.*, 2006; Yuill & Oakhill, 1988), our design focused on asking children inferencing questions while they listened to stories. It also closely followed van Kleeck's (2008) evidence-based recommendations for fostering inferential language in younger children, e.g., targeted questions and scripted feedback in a shared reading context.

Our findings have several important implications for the field. First, having used gold-standard methods to test the efficacy of supporting preschoolers' inferencing skills using implicit methods at home, the evidence base for this type of training remains negligible. Future interventions should offer more support for children of this age by using direct teaching methods, and should scrutinise the potential benefits of professionally implemented interventions that use explicit, well-defined, comprehension-focused activities. Second, our findings highlight the link between inferencing and general language ability. Specifically, we would like to promote strategies that strengthen vocabulary to provide a solid foundation for inferencing. The results of this rigorous RCT will benefit researchers engaged in theory-building and testing as well as practitioners choosing how to allocate resources.

Quality of parent delivery in the intervention was high. Videos of intervention sessions with pilot caregiver–child dyads showed that the training was accessible and implemented as intended. We designed the training to be consistent and easy to follow, and gave clear instructions in the support materials. Qualitative and quantitative comments from the intervention diaries and exit questionnaires showed that engagement was generally good, with a mean dosage of 24 sessions out of a recommended 20 over the month-long intervention. Responses to the Home Life questionnaire revealed that, for the vast majority of our inferencing training group, reading is a frequent and enjoyable activity. Ninety-five percent of returned questionnaires (n = 44) stated that someone reads or looks at books with their child daily (84%) or more than three times per week (11%), and 93% of parents who returned questionnaires agreed or strongly agreed that they found reading on their own enjoyable. While these caregivers are not necessarily representative of the general population, it suggests the format of the training was familiar and pleasurable, raising the likelihood of good quality implementation.

In addition to sound theoretical foundations and good treatment fidelity (according to our measures), our study used an RCT as the gold standard for testing the effectiveness of an intervention. Despite having used these three core strategies for maximising success, we are left with the question of why the training did not have reliable effects on our primary outcome of inferencing skills.

Recall that our original aims were to investigate: (i) whether training inferencing in the preschool years is possible; and (ii) if so, whether it is possible in this age group using implicit methods. Our results suggest not on both counts, at least in an intervention of this length. First, the children that our intervention targeted may not be developmentally ready to benefit from this kind of implicit inferencing training (where a parent asks an open question with basic scaffolding in the case of incorrect responses). This is supported by another intervention study with preschoolers, which found that, although mothers' inferential *yes/no* questions and statements predicted children's receptive vocabulary growth over six months, mothers' inferential *wh*-questions did not (Tompkins, Bengochea, Nicol, & Justice, 2017). The authors suggest that since open-ended *wh*-questions (similar to those used in our study) do not provide the child with the correct information (in contrast to closed questions and statements), preschoolers may need inferences to be made more explicit for them to facilitate language development (see also Carmiol, Matthews, & Rodríguez-Villagra, 2018).

Our approach was novel in its focus on the oral language of children in the preschool years. On the whole, comparable successful interventions have targeted children ranging from six to nine years old (Clarke *et al.*, 2010; McGee & Johnson, 2003; Yuill & Oakhill, 1988), due to their focus on reading comprehension as an outcome measure. Although inferences are within reach of children from three to four years old (Filiatrault-Veilleux *et al.*, 2016, and as shown by the distributions of scores on our tests of inferencing), evidence that the same skills can be trained in preschoolers is scant. To the best of our knowledge, a single study has shown that five-year-old children showed improvements in oral comprehension (including inferencing) after explicit training activities that spanned seven months (Bianco *et al.*, 2010). Crucially, to be effective, the training in that study had to comprise explicit, well-defined, comprehension-focused activities, i.e., not shared reading and discussion alone – a point we will return to below. Therefore, despite showing competence in inferencing and engaging with the training material, under-fives may not be able to transfer the skills they practised during the shared reading activities to a test situation.

The reasons for this apparent age threshold cannot be conclusively answered by our data, but one potential factor could be four-year-olds' immature executive function skills. The working memory (WM) demands of the inferencing task may have prevented children from responding even if in principle they could make relevant inferences. While some of the vignettes were presented in two halves and with picture prompts (Rover the Dog; Jessie's Birthday Party), others were presented without a break and without visual support (Birthday; A New Pet). The latter two vignettes were therefore quite long (211 and 161 words, respectively), and were also administered later in the session so fatigue effects are likely to have been at play. The lower mean scores for these particular vignettes relative to other vignettes administered at the same time-point suggest that WM demands may have impeded children's inferencing performance (see also Freed & Cain, 2017, for evidence that younger children benefit from a segmented format when being tested on inference-making). While many real-world inferences necessitate the retention in memory of large blocks of texts, future studies might explore reducing these demands with preschool children.

Returning to the second consideration of whether inference training is possible in preschoolers using implicit methods, the indirect nature of the instruction provided may also explain the null results. Our training was focused on parents asking inferencing questions and children answering them, with parents responding to incorrect answers using minimal, prompted, item-specific feedback. Although the shared reading materials were designed to highlight gaps in the text, relying on children to realise that these gaps existed and then make the required inference without more explicit feedback, may have overestimated their capabilities at this age: learning opportunities may have been too subtle to effect the hypothesised change. One reason for adopting this implicit approach was that explicit instruction is not easily accessible by four-year-olds, yet to begin formal education. Another was the challenge of training parents in explicit methods. The current evidence base for the effectiveness of explicit parent-delivered interventions is small (Burgoyne *et al.*, 2018; Huat See & Gorard, 2013) relative to the more substantial literature on the success of interventions by trained professionals (Bianco *et al.*, 2010; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; Fricke *et al.*, 2017; Rogde, Melby Lervag, & Lervag, 2016). Thus, inferencing interventions may be more effective if delivered by early years professionals who could adapt some of the explicit methods used in the classroom with older children, e.g., giving practice in text prediction or in lexical inference (McGee & Johnson, 2003; Yuill & Oakhill, 1988). Indeed, a single study has shown that these explicit approaches can be effective for four-year-olds' inferencing abilities over a longer period (Bianco *et al.*, 2010). This is not to say that the medium of shared book-reading is problematic in itself; explicit feedback can be integrated into natural book-sharing interactions via adult modelling (van Kleeck, 2008, p. 638). Indeed, the discussions between caregiver and child resulting from adult feedback is likely to be beneficial for inferencing training, compared to simply answering comprehension questions (Yuill & Oakhill, 1988).

Our results raise the more general question of whether it makes sense to train inferencing as an isolated skill in preschoolers, or to instead concentrate on other aspects of language such as vocabulary. We found that both baseline and post-test measures of inferencing were highly correlated with children's baseline language ability as measured by the CELF Language Content Index (tapping vocabulary breadth, concept development, sentence comprehension, and comprehension of lexical relationships). This suggests that our measure of inferencing – and indeed inferencing ability in general – might reflect general language ability, and we would welcome studies that further analyse the nature of this association. Language skill (or more specifically the vocabulary component) may be a more powerful determinant of inferencing ability than the type of inferencing training we administered. This explanation is in line with evidence from individual differences and longitudinal studies showing vocabulary knowledge to be a key predictor of inferencing (Currie & Cain, 2015; Language and Reading Research Consortium, Currie, & Muijselaar, 2019; Lucas & Norbury, 2015; Silva & Cain, 2015), and is also supported by the lexical quality hypothesis, which predicts that more precise knowledge of words promotes efficient text comprehension (Perfetti, 2007). If a child doesn't yet know a word (i.e., developing vocabulary breadth) or have a sufficiently rich representation of its meaning (i.e., developing vocabulary depth), they are less likely to integrate the word into the situation model to make the required inference during comprehension. In the case of semantic inferences, for example, a rich and robust knowledge of word meanings is required to map between a word and its synonyms, co-hyponyms, or superordinates (e.g., knowing that apples and bananas are types of fruit), thus greater vocabulary depth and richer semantic networks facilitate more efficient and more complex semantic inferences. In a recent study that analysed the concurrent and longitudinal relations between inference-making, vocabulary, and verbal working memory in four- to five-year-olds to eight- to nine-year-olds, both vocabulary breadth and (to a lesser extent) depth explained inference-making skill in the early grades, i.e., at the same age as our sample (Language and Reading Research Consortium *et al.*, 2019).

Furthermore, vocabulary can boost the memory processes recruited during inferencing. Robust word representations can support the maintenance of semantic information in verbal working memory (Nation, Adams, Bowyer-Crane, & Snowling, 1999; Walker & Hulme, 1999), and efficient maintenance of word meaning is necessary for integrating information distributed throughout the discourse. Thus, good vocabulary supports inferencing in (at least) two distinct ways. A more effective approach to scaffolding inferencing might be to focus on boosting vocabulary breadth, depth, and conceptual knowledge, all of which can be used in making inferences. As vocabulary has been implicated in the development of multiple aspects of comprehension (Marulis & Neuman, 2010, 2013; Ouellette, 2006; Wright & Cervetti, 2017), training word learning could be a powerful tool to benefit language across the board. Accruing a greater vocabulary size and a richer knowledge of word meaning may be of particular benefit to inferencing. Interestingly, there is emerging evidence for a reciprocal relationship between vocabulary and inferencing skill (Language and Reading Research Consortium *et al.*, 2019). This highlights the importance of practising both skills in the classroom to benefit not only the discrete skills, but also the ways that each can support the other.

Another reason for favouring a focus on lower-level language skills such as vocabulary and lexical relationships is the apparent lack of transfer in our data. That is, the lack of transfer between inferencing skills practised during the training to those required at post-test could be taken to suggest a lack of generalisability between semantic domains. For example, if a child can make the inference that sitting on sand means being at the beach, this does not guarantee that they can make a different type of inference, say about a character's motivation or the consequences of their actions. Although learning about the sand/beach connection would increase a child's knowledge about that specific domain, it may not be useful for higher-level, general-purpose inferencing ability (if such a thing exists). This hypothesis is also in line with the modest correlations between baseline and post-tests of inferencing ($r = 0.59$ in the control group and $r = 0.4$ in the training group). That is, it may not be possible to train 'general purpose' inferencing. If this is the case then a more fruitful approach to boosting inferencing may be to focus on vocabulary and the development of richly connected semantic networks.

Summary and recommendations

There are several reasons – separately or in combination – which might explain why our training was not effective in improving inferencing. Age of the children, use of implicit-exposure training, lack of transfer between inferencing domains, and the tentative link between inferencing and underlying language ability could all have limited its potential to effect change. Nonetheless, due to the firm evidence base suggesting that the development of inferencing can be supported (albeit in older children and/or using more explicit methods), and our use of a robust RCT, we had good reason to believe that the children might learn from the training. The fact that they did not means that the evidence for inferencing training using implicit methods with younger children remains negligible. Future interventions would need to offer more support for children of this age. Our findings suggest the following priorities for future research and practice.

First, the association that we found between language and inferencing skills suggests that a more effective approach to scaffolding inferencing in the preschool years might be to focus on promoting vocabulary to develop the broad, deep, and rapidly accessed semantic knowledge necessary to make inferences viable. This should be preceded by in-depth analysis of the links between inferencing and language ability.

Second, interventions using more explicit inferencing training could shed light on whether the implicit nature of the current intervention was the limiting factor for preschoolers. To do this, materials could be adapted from successful interventions for older children (e.g., Bianco *et al.*, 2010; McGee & Johnson, 2003; Yuill & Oakhill, 1988) to determine whether explicit inferencing training could be accessible and effective for preschoolers. It may be that this type of training is best delivered professionally, i.e., at preschool. As reviewed above, the most successful inferencing interventions have used direct teaching methods, with frequent, explicit focus on the target skill or structures, and have often used group-based delivery methods (Elleman, 2017). These specialist skills that might allow children to gain insights into inference-making are likely best found in well-trained teachers.

Third, we would support future lab-based experimental studies to unpick the components of specific types of inferences, the inferencing-making process, and the associated cognitive resources (e.g., working memory, background knowledge, vocabulary – including its speed of access; Cain, Lemmon, & Oakhill, 2004; Freed & Cain, 2017; Language and Reading Research Consortium, 2018; Oakhill, Cain, & McCarthy, 2015). While the current study purposely chose to test a wide range of inference types, future studies might select from a more restricted range to investigate how different inference types vary in their developmental trajectories, how responsive they are to training, and how they are underpinned by different cognitive resources. This kind of research programme would provide new insights into limiting factors in early development and how best to support inference development at different points of development. This knowledge could also inform more effective tests for assessing inferencing in the preschool years, e.g., exploring the use of graphic organisers to support memory demands and organise key ideas when inferencing (Nesbit & Adesope, 2006), and using the most supportive dyadic contexts for inferencing, e.g., allowing children to demonstrate their inferencing abilities via narratives rather than undergoing formal question-and-answer tests (van Kleeck, 2006, p. 292). There is also scope for future studies to adopt a joint enquiry approach in which parents and children collaboratively answer questions, giving parents the opportunity to be reading role models and model their own inferencing and deduction processes.

Although our intervention was designed to maximise fidelity – indeed we have no reason to infer that parents did not administer it as intended – future studies should monitor implementation directly, e.g., by asking families to record their intervention sessions and then analysing a proportion of these against the protocols (e.g., Noble *et al.*, unpublished observations). In studies which find no effect of the chosen intervention, direct monitoring would provide specific information about why parent-delivered interventions are not effective. Prior to further RCTs, feasibility studies are essential for clarifying the acceptability of proposed interventions to stakeholders.

Finally, we would like to reiterate the importance of reporting and publishing null results. Our findings show the usefulness of rigorously evaluating well-founded interventions to inform future work, and of disseminating the findings to practitioners who can use the emerging evidence in identifying and developing effective practices.

Acknowledgements. This research was supported by ESRC grant ES/M003752/1. The authors thank the families for their participation, and Kiera Solaiman, Charlotte Rowley, Annalise Guild, Lauren Lofthouse, Katie Gascoigne, Hannah Putrus, Rachael Staunton, and Jonathan Turner for assistance with collecting and coding data. They also thank Dr Tim Heaton for conducting the randomisation, Dr Jamie Lingwood for assistance with CLAN, Dr Paula Clarke for comments on an earlier draft, and members of the shared book-reading grant team for advice and support. CR is also supported by the ESRC International Centre for Language and Communicative Development (LuCiD; ES/L008955/1). The authors have declared that they have no competing or potential conflicts of interest in relation to this paper.

References

Ackland, J. (2012). *At home with counting.* Oxford University Press.

Alcock, K. J., Meints, K., & Rowland, C. F. (2017). UK-CDI Words and Gestures – Preliminary norms and manual. Retrieved from <http://lucid.ac.uk/ukcdi>.

Anderson, A., Anderson, J., Lynch, J., Shapiro, J., & Kim, J. (2012). Extra-textual talk in shared book reading: a focus on questioning. *Early Child Development and Care*, *182*, 1139–54.

Ard, L. M., & Beverly, B. L. (2004). Preschool word learning during joint book reading: effect of adult questions and comments. *Communication Disorders Quarterly*, *26*(1), 17–28.

Baker, N. D., & Nelson, K. E. (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. *First Language*, *5*, 3–21.

Bianco, M., Bressoux, P., Doyen, A. L., Lambert, E., Lima, L., Pellenq, C., & Zorman, M. (2010). Early training of oral comprehension and phonological skills at preschool: the results of a 3 year longitudinal study. *Scientific Studies of Reading*, *14*(3), 211–46.

Bos, L. T., De Koning, B. B., Wassenburg, S. I., & van der Schoot, M. (2016) Training inference making skills using a situation model approach improves reading comprehension. *Frontiers in Psychology*, *7*, e00116.

Burgess, S. (2010) The role of shared reading in the development of phonological awareness: a longitudinal study of middle to upper class children. *Early Child Development and Care*, *127*, 191–99.

Burgoyne, K., Gardner, R., Whiteley, H., Snowling, M. J., & Hulme, C. (2018). Evaluation of a parent-delivered early language enrichment programme: evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *59*(5), 545–55.

Bus, A. G., Leseman, P. P., & 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*, 53–76.

Butterworth, N. (2011) *One snowy night (a tale from Percy's Park)*. London: HarperCollins.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489–503.

Cain, K., Lemmon, K., & Oakhill, J. (2004). Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–81.

Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and child directed speech. *First Language*, *33*, 268–79.

Carmiol, A. M., Matthews, D., & Rodríguez-Villagra, O. A. (2018). How children learn to produce appropriate referring expressions in narratives: the role of clarification requests and modeling. *Journal of Child Language*, *45*(3), 736–52.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: a randomised controlled trial. *Psychological Science*, *21*(8), 1106–16.

Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & Fey, M. E. (2015). The efficacy of recasts in language intervention: a systematic review and meta-analysis. *American Journal of Speech-Language Pathology*, *24*, 237–55.

Colmar, S. (2014). A parent-based book-reading intervention for disadvantaged children with language difficulties. *Child Language Teaching and Therapy*, *30*, 79–90.

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004). Beyond the reading wars: exploring the effect of child–instruction interaction on growth in early reading. *Scientific Studies of Reading*, *8*, 305–36.

Currie, N. K., & Cain, K. (2015). Children's inference generation: the role of vocabulary and working memory. *Journal of Experimental Child Psychology*, *137*, 57–75.

Das Gupta, P., & Bryant, P.E. (1989) Young children's causal inferences. *Child Development*, *60*, 1138–46.

Dawes, E., Leitão, S., Claessen, M., & Kane, R. (2019). A randomized controlled trial of an oral inferential comprehension intervention for young children with developmental language disorder. *Child Language Teaching and Therapy*, *35*(1), 39–54.

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: a meta-analytic review. *Journal of Educational Psychology*, *109*, 761–82.

Ezell, H. K., & Justice, L. M. (2000). Increasing the print focus of adult–child shared book reading through observational learning. *American Journal of Speech-Language Pathology*, *9*(1), 36–47.

Farrar, M. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language*, *17*, 607–24.

Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2016). Comprehension of inferences in a narrative in 3- to 6-year-old children. *Journal of Speech, Language, and Hearing Research*, *59*, 1099–110.

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: the role of verbal and inferential skills. *Discourse Processes*, *48*, 119–38.

Freed, J., & Cain, K. E. (2017), Assessing school-aged children's inference making: the effect of test story format in listening comprehension. *International Journal of Language and Communication Disorders*, *52* (1), 95–105.

Fricke, S., Bowyer-Crane, C., Haley, A., Hulme, C., & Snowling, M. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *53*, 280–90.

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., Snowling, M., & Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *58*, 1141–51.

Girolametto, L., & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools*, *33*, 268–81.

Hammett, L. A., van Kleeck, A. V., & Huberty, C. J. (2003). Patterns of parents' extratextual interactions during book sharing with preschool children: a cluster analysis study. *Reading Research Quarterly*, *38*, 442–68.

Heath, S. B. (1983). *Ways with words: language, life, and work in communities and classrooms*. New York: Cambridge University Press.

Hindman, A. H., Connor, C. M., Jewkes, A. M., & Morrison, F. J. (2008). Untangling the effects of shared book reading: multiple factors and their associations with preschool literacy outcomes. *Early Childhood Research Quarterly*, *23*(3), 330–50.

Hoff-Ginsberg, E. (1991). Mother–child conversation in different social classes and communicative settings. *Child Development*, *62*, 782–96.

Huat See, B., & Gorard, S. (2013). *What do rigorous evaluations tell us about the most promising parental involvement interventions? A critical review of what works for disadvantaged children in different age groups*. London: Nuffield Foundation.

Hulme, C., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M. J. (2009). *York assessment of reading for comprehension: early reading*. London: GL Assessment.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*, 343–65.

Karweit, N. (1989). The effects of a story-reading program on the vocabulary and story comprehension skills of disadvantaged preschool and kindergarten students. *Early Education and Development*, *1*, 105–14.

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: inquiry, instruction, and social interaction. *Review of Educational Research*, *67*, 271–99.

Lakens, D., McLatchie, N., Isager, P. M., Scheel, A. M., & Dienes, Z. (2018). Improving inferences about null effects with Bayes factors and equivalence tests. Journals of Gerontology: Series B, gby065.

Lakens, D., Scheel, A. M., & Isager, P. (2018). Equivalence testing for psychological research: a tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–69.

Language and Reading Research Consortium (2015). The dimensionality of language ability in young children. *Child Development*, *86*(6), 1948–65.

Language and Reading Research Consortium (2017). Pressure points in reading comprehension: a quantile multiple regression analysis. *Journal of Educational Psychology*, *109*(4), 451–64.

Language and Reading Research Consortium (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading*, *22*(2), 117–36.

Language and Reading Research Consortium, Currie, N. K., & Muijselaar, M. M. L. (2019). Inference making in young children: the concurrent and longitudinal contributions of verbal working memory and vocabulary. Journal of Educational Psychology. https://doi.org/10.1037/edu0000342

Law, J., Charlton, J., Dockrell, J., Gascoigne, M., McKean, C., & Theakston, A. (2017). *Early language development: needs, provision and intervention for pre-school children from socio-economically disadvantaged backgrounds*. London: Education Endowment Foundation.

Lucas, R., & Norbury, C. (2015). Making inferences from text: it's vocabulary that matters. *Journal of Speech, Language, and Hearing Research*, *58*, 1224–32.

MacWhinney, B. (2000). *The CHILDES Project: tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Review of Educational Research*, *80*, 300–35.

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young at-risk children's word learning: a meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*, 224–63.

McGee, A., & Johnson, H. (2003). The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology*, *23*(1), 49–59.

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. *Journal of Child Psychology and Psychiatry*, *58*(10), 1122–31.

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*, 7–26.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, *73*, 139–58.

National Foundation for Educational Research (2015). *NFER Reception Baseline Assessment*. Slough: NFER.

Neale, M. D. (1989). *Neale Analysis of Reading Ability – Revised*. Windsor: NFER-Nelson.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: a meta-analysis. *Review of Educational Research*, *76*(3), 413–48.

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C. F., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: a meta-analysis. *Educational Research Review*, *28*. https://doi.org/10.1016/j.edurev.2019.100290

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*(1), 31–39.

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121.

Oakhill, J., Cain, K., & McCarthy, D. (2015). Inference processing in children: the contributions of depth and breadth of vocabulary knowledge. In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 140–59). Cambridge University Press.

Ouellette, G. P. (2006). What's meaning got to do with it? The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*, 554–66.

Oxford University Press (2018). *Why closing the word gap matters: Oxford language report*. Oxford University Press. Retrieved from <oxford.ly/wordgap>.

Pellegrini, A. D., Perlmutter, J. C., Galda, L., & Brody, G. H. (1990). Joint reading between black Head Start children and their mothers. *Child Development*, *61*, 443–53.

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–83.

Potter, C. A., & Haynes, W. O. (2000). The effects of genre on mother–toddler interaction during joint book reading. *Infant–Toddler Intervention*, *10*, 97–105.

Price, L. H., van Kleeck, A., & Huberty, C. J. (2009). Talk during book sharing between parents and preschool children: a comparison between storybook and expository book conditions. *Reading Research Quarterly*, *44*, 171–94.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: a visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*(1), 115–29.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org>.

Rogde, K., Melby-Lervag, M., & Lervag, A. (2016). Improving the general language skills of second-language learners in kindergarten: a randomized controlled trial. *Journal of Research on Educational Effectiveness*, *9*, 150–70.

Rowe, M. (2013). Decontextualized language input and preschoolers' vocabulary development. *Seminars in Speech and Language*, *34*(4), 260–6.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. Online <http://www.rstudio.com/>.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, *26*(6), 515–20.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*, 18. doi:10.1186/1741-7015-8-18

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-year-old children make relevance inferences in indirect verbal communication. *Child Development*, *84*(6), 2079–93.

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, *107*, 321–31.

Snow, C. E. (1993). Families as social contexts for literacy development. In W. Damon (Series Ed.) & C. Daiute (Vol. Ed.), *New directions in child development: Vol. 61. The development of literacy through social interaction* (pp. 11–24). San Francisco, CA: Jossey-Bass.

Snow, C. E. (2001). *Reading for understanding*. Santa Monica, CA: RAND Education & the Science and Technology Policy Institute.

Tompkins, V., Bengochea, A., Nicol, S., & Justice, L. M. (2017). Maternal inferential input and children's language skills. *Reading Research Quarterly*, *52*(4), 397–416.

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*, 1106–14.

van Kleeck, A. (2006). Fostering inferential language during book sharing with prereaders: a foundation for later text comprehension strategies. In A. van Kleeck (Ed.), *Sharing books and stories to promote language and literacy* (pp. 269–318). San Diego, CA: Plural Publishing.

van Kleeck, A. (2008). Providing preschool foundations for later reading comprehension: the importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools*, *45*, 627–43.

van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in head start preschoolers with language impairment using scripted book-sharing discussions. *American Journal of Speech-Language Pathology*, *15*(1), 85–95.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1256–71.

Wiig, E., Second, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals Preschool 2 UK*. London: Pearson.

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, *52*(2), 203–26.

Xiao, Z., Kasim, A., & Higgins, S.E. (2016) Same difference? Understanding variation in the estimation of effect sizes from educational trials. *International Journal of Educational Research*, *77*, 1–14.

Yanez Diaz Barriga, A. (2018). *Domain-general precursors of children's mathematics skills: the role of working memory and language* (Unpublished doctoral dissertation). University of Sheffield, UK. Online <http://etheses.whiterose.ac.uk/23822/>..

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading-comprehension. *Applied Cognitive Psychology*, *2*(1), 33–45.

Zucker, T. A., Justice, L. M., Piasta, S. B., & Kaderavek, J. N. (2010). Preschool teachers' literal and inferential questions and children's responses during whole-class shared reading. *Early Childhood Research Quarterly*, *25*(1), 65–83.

Appendices[2]

Appendix A: Trial registration

Appendix B: CONSORT 2010 checklist

Appendix C: Inferencing training video script

Appendix D: Inferencing intervention support leaflet

Appendix E: Inferencing intervention materials: book titles, inference-eliciting questions, and model responses

Appendix F: Inferencing test materials: bespoke inferencing vignettes, comprehension questions, and coding scheme

Appendix G: Scores by inference question and inference type

[2]Available at <https://osf.io/95qr8>.

Cite this article: Davies C, McGillion M, Rowland C, Matthews D (2020). Can inferencing be trained in preschoolers using shared book-reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability. *Journal of Child Language* **47**, 655–679. https://doi.org/10.1017/S0305000919000801