An Evidence-Based Approach to Teach Inferential Language During Interactive Storybook Reading With Young Children

Elizabeth Spencer Kelley
University of Missouri - Columbia
Cite this document as:
Structured Abstract

Clinical Question: Would young children demonstrate improvements in inferential question-answering after interactive book-reading intervention that targeted inferential questions in comparison to a similar intervention that targeted overall language ability?

Method: Systematic Review

Study Sources: ERIC, Academic Search Complete, ASHAWire

Search Terms: inferential questions OR inferences OR inferential language AND intervention OR instruction OR treatment OR language impairment

Number of Included Studies: 3

Primary Results:

Children who participated in interactive reading interventions that targeted inferential language, including inferential question-answering, demonstrated improvements in literal and inferential language.

No direct comparisons of interventions to target inferential language and interventions to target overall language were available.

Conclusions: Although there is substantial evidence to support interactive book-reading as an effective strategy for language learning in young children, few studies have examined teaching inferential language in this context. The small number of studies available suggests that strategies such as prompting for responses to questions, modeling appropriate responses, and including “think aloud” explanations of responses may be appropriate for teaching inferential question-answering in young children. There is a need for additional research to identify the best approaches to teach inferential language, including answering inferential questions about stories, to young children.
Clinical Scenario

Kate is a speech-language pathologist who works in an elementary school serving many children from low-income families. As part of her practice, she works with kindergarten children, many who have limited oral language skills but have not been identified with language impairment. For these children, her primary service delivery model is classroom-based biweekly small-group sessions that focus on language skills in the context of interactive book-reading activities. Kate uses an interactive reading style to provide multiple opportunities for children to engage with the book and respond to questions. Kate bases her interactive reading approach on dialogic reading (Whitehurst et al., 1994). She selects age-appropriate books, including some of the same books that are used in the whole-group classroom instruction provided by the teacher. During the small-group sessions, Kate reads the storybook aloud, embeds multiple opportunities for children to respond to questions and make comments about the story, and uses recasts and expansions to teach vocabulary and other language targets. As part of the interactive book-reading, Kate targeted two primary types of story questions: literal questions and inferential questions. Literal questions are questions that the child can answer by recalling information from the story. For example, Kate asks children to answer questions like, *Who was in this story? What was Ashley doing? What happened at the end of the story?* For most literal questions, the answer to the question is directly stated in the story or shown in an illustration.

Kate also targeted inferential questions during her small-group sessions. Inferential questions require the child to make an inference about a character or event in the story. The child might be asked to make an inference about a character’s emotion, talk about the reasons for a character’s action, or make a prediction about events in the story. The necessary information to answer inferential questions is not usually directly stated in the text or shown in the pictures. Inferential questions are generally more abstract than literal questions and, often, the child needs to make a connection between the story and his or her own knowledge. Examples of inferential questions include, *Why was Anna happy? Why did Theo help Ben? and What do you think will happen next?* At the end of the first semester, Kate carefully examines the data she collected during sessions and from brief progress-monitoring assessments. During sessions, Kate kept a quick tally of correct and incorrect responses to questions about the story for each child. Because she typically reads the same story repeatedly on consecutive days, she can monitor how each child is improving in his or her ability to answer the questions he or she has been practicing with the same story. In general, she notices that children improve across sessions with the same storybook, frequently learning to answer most of the literal questions and many of the inferential questions by the end of several readings. Kate conducted three brief progress-monitoring assessments during the semester. For these assessments, Kate reads a new story with each child individually and asks the child to respond to literal and inferential questions about the story. Because the children have not heard the stories or the questions before, progress-monitoring sessions tell Kate how children are improving in their overall ability to answer questions about stories. The data from the progress-monitoring sessions indicates that the children improved their ability to answer literal questions about stories. For example, one child improved from 30% correct at the first progress-monitoring assessment to 90% correct at the most recent time point. However, the data indicates that children still struggle with inferential questions. The same child averaged around 30 to 40% correct at all three progress-monitoring sessions.

During classroom reading activities, Kate observed that other children in the kindergarten class are able to answer these types of inferential questions. She believes that answering questions about stories is an important skill for the children she works with, because she knows that early language skills contribute to later reading comprehension ability. Reflecting on her interactive book-reading sessions, Kate noted that she frequently spent more time on, and provided more practice opportunities for, literal questions. She thought that increasing the focus on inferential questions might be one strategy to improve the kindergarten
children’s ability to answer inferential questions. However, Kate hoped that she could find evidence to support other specific practices.

**Background Information**

As part of a multi-tiered system of support (MTSS or Response to Intervention model), tiers of services are delivered to support the learning needs of all children. In many schools, one of the newer roles for the speech-language pathologist is to contribute to efforts to prevent later academic difficulties (Ehren, Montgomery, Rudebusch, & Whitmire, 2006). Special-education funds can be devoted to support children who may need additional support (i.e., Tier Two), but who have not yet been identified as needing special-education services. Kate is currently in this position. She collaborates with the kindergarten teacher to work with a group of children who do not meet benchmarks in the oral language area. Her goal is to support the learning needs of these children sufficiently so they make adequate progress.

**Interactive Book-Reading Intervention**

Interactive book-reading is an evidence-based practice with strong research support. In a meta-analyses of 31 studies of interactive book-reading with preschool and kindergarten children, Mol, Bus, and de Jong (2009) reported moderate effects on expressive and receptive vocabulary (overall $d = .54$). Interactive book-reading programs have been demonstrated to improve vocabulary knowledge of young children when administered individually or in small groups, implemented by parents or teachers, and with groups of children who are at risk due to socioeconomic factors or limited oral vocabulary (Dale, Crain-Thoreson, Notari-Syverson, & Cole, 1996; Hargrave & Sénéchal, 2000; Lonigan, Anthony, Bloomfield, Dyer, & Samwel, 1999; Whitehurst et al., 1994; Whitehurst et al., 1988). Although many studies of interactive book-reading have focused on vocabulary knowledge as an outcome, few studies have examined interactive book-reading as an approach to improve other oral language skills.

**Inferential Language**

Kate’s interest in the inferential language abilities of the children she works with is well warranted. In preschool-age children, the ability to make inferences is related to comprehension of narratives (Kendeou, Bohn-Gettler, White, & Van Den Broek, 2008; Lepola, Lynch, Laakonen, Silvén, & Niemi, 2012) and in school-age children, the ability to make inferences is predictive of later reading comprehension (Cain, Oakhill, & Lemmon, 2004). For the children that Kate works with, her observations of limited inferential language are not surprising. Children with language difficulties are likely to struggle with both literal and inferential language (Bishop & Adams, 1992) and there is evidence inferential language may be particularly challenging (Blank, Rose, & Berlin, 2003). Children with language impairment perform poorly relative to typically developing peers on inferential tasks, and performance on these tasks is related to language comprehension (Adams, Clarke, & Haynes, 2009; Ford & Milosky, 2003). Difficulties with inferential language and deficits in reading comprehension are interrelated (Cain et al., 2004; Cain & Oakhill, 1999).

In studies of parent–child interactions during storybook sharing activities, children who were exposed to more inferential language used more of it and improved in inferential language abilities (van Kleeck, Gillam, Hamilton, & McGrath, 1997). Importantly, children who participated in storybook sharing with inferential language had higher scores on measures of reading comprehension in the third grade than peers who were less engaged in storybook sharing (Serpell, Baker, & Sonnenschein, 2005). Van Kleeck (2008) argues that inference-making contributes to later text comprehension by encouraging children to make connections between information in the text and their own knowledge. Indeed, text comprehension strategies taught to older children frequently include strategies for generating inferences and making connections with background knowledge (see for a review: Rosenshine & Meister, 1994).

**Clinical Question**

Kate used the PICO framework to develop a targeted, clinically-relevant research question, as suggested by the American Speech-Language-Hearing Association (ASHA). She defined (P) the population, (I) the intervention, (C) the comparison intervention, and (O) the intended outcome as:

- **P** – young children with or without language difficulties
- **I** – interactive book-reading intervention with a focus on inferential questions
- **C** – interactive storybook reading with a more general language focus

Copyright © 2015 NCS Pearson, Inc. All rights reserved.
Kate’s question was: Would young children demonstrate improvements in inferential question-answering after interactive book-reading intervention that targeted inferential questions in comparison to an interactive book-reading intervention that targeted overall language ability?

Search for the Evidence

Inclusion Criteria

Before she began to look for articles, Kate thought carefully about the criteria for her search. She wanted to find high-quality evidence so she decided to limit her search to articles published in peer-reviewed journals. She would include studies that used a number of study designs (experimental, quasi-experimental, single-case design). Although she was most interested in studies that included children around the kindergarten age, she thought that studies that included preschool-age or early school-age children could provide important information because the language abilities and intervention approaches appropriate for children in this age range would be similar. She decided that her primary inclusionary criteria would be (a) a published, peer-reviewed study of; (b) an interactive book-reading intervention that targeted inferential questions with; (c) preschool-age or early elementary-age children as participants; and (d) a measure of inferential language included as an outcome.

Kate began her search using two databases: ERIC and Academic Search Complete. Figure 1 provides details of the search process.

She conducted multiple searches using combinations of the keywords inferential questions, inferences, inferential language, AND intervention, instruction, treatment. Kate found that searches using more general search terms like story comprehension AND questions resulted in too many unrelated articles (638 articles in ERIC). Although Kate did not limit her search to articles focused on children with language difficulties, she also conducted searches with language impairment in combination with her other search terms. These searches resulted in over 160 articles, including many duplicates. Kate continued her search using the ASHAWire database but did not identify any additional articles. Her next step was to review the titles and abstracts to identify articles that were appropriate. Many of the articles identified in the search did not meet criteria (i.e., were not treatment articles, did not specifically target inferential language). Kate next reviewed the full text of the most relevant articles and searched the reference list of each of those articles to see if any additional articles were listed and located one additional article. During her search, she also located one review article (Hall, 2015) but no articles included in that paper met her search criteria.

Evaluating the Evidence

Summaries of Included Studies

Three studies met Kate’s original search criteria: van Kleeck et al. (2006), Desmarais, Nadeau, Trudeau, Filiatrault-Veilleux, & Maxès-Fournier (2013), and Bradshaw, Hoffman, and Norris (1998). See Table 1 for summaries of the three studies.

Van Kleeck et al. (2006) delivered an individualized interactive-reading intervention to 30 preschool children with language impairments. Participants attended Head Start centers and were randomly assigned to either the intervention condition or a no-intervention control group. In the intervention condition, children participated in brief (15 minute) sessions twice a week for eight weeks. Trained research assistants led children through scripted interactive-reading activities using storybooks that had been modified for this purpose. The two storybooks, Mooncake (1983) and Skyfire (1984) by Frank Asch were selected for similarities in complexity and length. Three scripts for each book were developed based on previous study of parents and children reading the same two books (van Kleeck et al., 1997). Each script included 25 questions, with 25% of them inferential. Samples from a script were included in an appendix (p. 95).

For each question, the script included opportunities for the child to respond and prompts to guide the child to a correct response. For example, for the question How do you think Bear feels because his friend Little Bird is leaving?, a prompt might highlight key information in the story using a cloze procedure, Maybe he feels sad because he won’t see (point to picture of Little Bird) _______. When children responded appropriately, the adult gave a confirmation, Yes. I think he feels sad because he won’t see his friend for a long time. When children did not respond appropriately, a correct response and explanation was modeled by the adult. For example, the adult might say, Bear is sad because he won’t see his friend for a long time. Do you ever feel sad when you won’t see someone for a long time? The authors described the modeled responses
and explanation as “thinking aloud,” a strategy that would help children understand the story and how to respond to questions.

At pre- and posttest, participants received a measure of receptive vocabulary, the Peabody Picture Vocabulary Test, Third Edition (PPVT™-III; Dunn & Dunn, 1997) and a measure of literal and inferential language, the Preschool Language Assessment Instrument (PLAI; Blank, Rose, & Berlin, 1978). The PLAI was scored to create two subscores, one for literal language and one for inferential language. Participants who received the intervention demonstrated greater gains than the control group in receptive vocabulary, literal language, and inferential language with medium to large effect sizes.

Desmarais et al. (2013) provided a similar intervention to 16 children (4- through 6-year-olds) with language impairments. The study used a multivariate repeated-measures design and there was no comparison group. Participants received 10 sessions (15 to 20 minutes each) of individual intervention provided by their speech-language pathologist. The authors stated that intervention procedures were modeled from van Kleeck et al. (2006) and consisted of interactive book-reading with scripted literal and inferential questions. Two books were selected for length, illustrations, and narratives structure. Each book-reading session included 16 questions, eight of which were inferential, with a hierarchy of prompts to guide children to appropriate responses. Prompts began with a simpler form of the question (i.e., rephrasing a question into a cloze statement), followed by a semantic cue (i.e., adding information from the story), and finally a phonemic cue (i.e., giving the first sound or syllable in the response). After prompting, children were asked to answer the question again. Outcome measures were the PLAI and a researcher-created measure of story comprehension. The PLAI was scored to provide one general score, rather than subscores for literal and inferential language. The researcher-created measure included a book-reading activity similar to the intervention procedures with 10 literal and 10 inferential questions. Two forms of the measure were created and were alternately administered at four time points: prebaseline, pretest, posttest, and delayed posttest. On the PLAI, the participant group mean increased from pretest to posttest. On the researcher-created measure, scores on the inferential questions increased across the four assessment points, but there was no statistically significant increase between pretest (39% correct on inferential questions) and posttest (43% correct). The authors explained that because two different forms of the assessment were used at pretest and posttest, results may have been due to differences in the difficulty level of the forms.

Bradshaw et al. (1998) used an alternating treatments single-case design to examine the effect of two interactive book-reading interventions. Participants were two 4-year-old boys with language delays. In one condition, expansions and cloze procedures were used to guide children to generate inferences. The adult requested a label from the child, expanded the label into a sentence or phrase, and then used a cloze procedure to generate an inference (e.g., The cow is taking a bath because…; p. 88). In the second condition, the adult asked literal and inferential questions and modeled responses to the questions (e.g., Why will he eat the rose? [pause for response] He will eat the rose because he is hungry; p. 88). Two books, Mrs. Wishy Washy (1990a) and The Red Rose (1990b) by Joy Cowley were selected for the study and each was assigned to one of the intervention conditions. These books were selected due to their series of events as opposed to a single main event. At each session, transcriptions of the participants’ responses were used to measure frequency of responses, frequency of interpretations (inferences), and frequency of syntactic forms. For both children, frequency of inferences was higher in the expansion and cloze intervention than in the question-answering condition.

Quality of Evidence

To evaluate the evidence provided by these three studies, Kate made use of the Critical Appraisal of Treatment Evidence (CATE) framework recommended by Dollaghan (2007). The CATE framework includes a set of questions to guide the evaluation of treatment studies. Kate prioritized a few questions related to study design, treatment implementation, and findings: Was the evidence from an experimental study? Was there a control group or condition? Was the treatment described clearly and implemented as intended? Was the finding statistically significant? and Was the finding important? (Dollaghan, 2007; p. 153).

Van Kleeck et al. (2006) utilized an experimental group design with random assignment to treatment and control conditions. The treatment was explained clearly and included information about treatment procedures, duration and intensity, and training of research assistants who provided the treatment. Examples of scripts were included in the appendix of the article. To monitor treatment implementation, videotapes of intervention sessions were
reviewed once per week. However, no report of fidelity was provided. Findings on the primary outcome measures, a standardized norm-referenced measure of receptive vocabulary (PPVT-III) and a standardized assessment with subscores for literal and inferential questions (PLAI), were statistically significant and effect sizes were medium to large. Because the study used a randomized control design, this study had the highest level of evidence.

Desmarais et al. (2013) used a multivariate repeated measures design with no control group. The explanation of the treatment was somewhat brief and there was only limited information about the treatment procedures. Information about dosage and intensity and some examples of questions and prompts were provided. However, there was no report of the training of the speech-language pathologists who administered the intervention and there was no report of treatment fidelity. Findings on the primary outcome measure, a researcher-created measure of literal and inferential language, were nonsignificant for the intervention period. The participants’ scores on a standardized measure of literal and inferential language (PLAI) were significantly higher at posttest than at pretest, but because there was no comparison group it was not possible to attribute this change to the intervention. The lack of comparison group or treatment meant that this study provided a weak level of evidence relative to van Kleeck et al. (2006).

Bradshaw et al. (1998) used an alternating treatments single-case design. Some of the questions in the CATE framework did not easily apply to the single-case design, so Kate applied the quality indicators for single-case design recommended by Horner et al. (2005). Related to study design, the study did not include baseline measurement of the target behavior, making it difficult to interpret performance during the two treatment conditions. Two participants were included in the study; this did not meet the minimum of three replications of effect recommended by Horner et al. The intervention conditions were described briefly but clearly. Although a section on intervention fidelity was included in the manuscript, the information provided in this section did not relate to treatment implementation. The authors reported that the frequency of inferences was higher in the expansion and cloze intervention condition than the question-answering condition. However, the two interventions were delivered with different books, meaning that differences in inference use might be due to the books rather than the intervention. Because the study had no baseline data and fewer than three replications of treatment effect, the level of evidence provided was also weak relative to van Kleeck et al. (2006).

The Evidence-Based Decision

After reviewing the studies, Kate revisited her PICO question: Would young children demonstrate improvements in inferential question-answering from an interactive reading intervention that targeted inferential language? Just one study, van Kleeck et al. (2006) had a strong study design with a control group. Participants in the intervention made gains in inferential language between pretest and posttest and the effect sizes for these gains were medium to large. The control group in van Kleeck et al. did not receive any intervention, so it was not possible to compare these results with another interactive reading intervention. However, the evidence from this study suggested to Kate that specifically targeting inferential questions during interactive book-reading and using prompts and modeled responses would be appropriate strategies to incorporate into her practice.

Because of issues with the study design (i.e., no comparison group, measurement of treatment outcomes), Kate did not feel that the results of Desmarais et al. (2013) could be used to support the use of the treatment strategies in that study. Although Desmarais et al. described the intervention as being modeled after the work of van Kleeck et al. (2006), Kate noticed some important differences. For example, Desmarais et al. describe a hierarchy of prompts (e.g., a phonological cue) that did not seem to support the inferencing task. The authors did not provide a rationale for the use of these prompts. Kate decided that the Desmarais et al. study would not be part of her evidence-based decision.

Bradshaw et al. (1998) used an alternating treatments single-case design. Although single-case designs can provide strong evidence of treatment effects (Horner et al., 2005), the lack of baseline data and fewer than three replications of treatment effect in this study were troubling. However, when Kate examined the graphs of results, it was clear to her that both participants demonstrated an increase in the ability to answer inferential questions. Kate decided that Bradshaw et al. provided preliminary evidence of an intervention procedure that might be useful.

At the beginning of her review, Kate hoped to find clear recommendations for an interactive book-reading treatment approach that would help her facilitate inferential question-answering in young children. She was disappointed to find that few studies met her search criteria. Of the studies
that she did find, descriptors of treatment procedures were somewhat limited, making it difficult to easily incorporate the techniques into her own practice. However, Kate felt that she learned valuable information related to teaching inferential language and decided to incorporate strategies from van Kleeck et al. (2006) and Bradshaw et al. (1998) into her interactive book-reading sessions.

Kate created scripts of questions and prompts for the storybooks she used with the kindergartners in the next few weeks. She selected a balance of both literal and inferential questions, using a ratio of 60% literal and 40% inferential. In van Kleeck et al. (2006), 25% of questions were inferential. Kate decided that because her students had already demonstrated progress on literal questions, she would increase the proportion of inferential questions to provide more practice opportunities. For each question, she included options for confirmations of appropriate answers (e.g., Yes. He is happy because he has a new bike!) as well as prompts and “think aloud” responses for occasions when children did not respond appropriately. Drawing on the Bradshaw et al. (1998) study, Kate chose prompts that included expansion and cloze procedures (e.g., She is worried because ____).

To design the “think aloud” responses described in van Kleeck et al. (2006), Kate referred to an article that she had identified during her search that did not meet her search criteria: a tutorial by van Kleeck (2008) that provided specific recommendations for practitioners targeting inferencing during interactive storybook reading. The strategies described in van Kleeck (2008) overlapped with those implemented in van Kleeck et al. (2006) and additional detail was provided about the “think aloud” strategy. The author explains that the “think aloud” strategy should provide a model of how to make an inference for children who aren’t yet able to do it on their own. Kate created “think alouds” that highlighted key information in the story or modeled connections with background knowledge to explain the rationale for the modeled response. For example, “think alouds” for inferential questions about character emotions (e.g., How do you think Danny feels?) included a connection to the child’s own experiences (e.g., I think he feels excited! I feel excited when I get to go to a birthday party. Do you?).

Kate incorporated these scripts into the interactive book-reading that she used to focus on other language targets including vocabulary and grammar. As before, Kate carefully monitored progress on both literal and inferential question-answering to determine if her treatment was effective. In addition to data collected during book-reading sessions and on brief progress-monitoring measures, Kate decided that she would occasionally record and transcribe children’s language during sessions to examine their responses to questions and prompts. This would allow her to evaluate and revise her scripts for future interactive book-reading sessions. Overall, Kate felt confident in her approach to improve inferential language skills in young children.

Author Note

Elizabeth Spencer Kelley, PhD, CCC-SLP is an assistant professor in the Department of Communication Science and Disorders, School of Health Professions, at the University of Missouri-Columbia. Her research focuses on the development of language and literacy skills of young children with a focus on children with limited oral language skills.

Elizabeth Kelley
Department of Communication Science and Disorders
University of Missouri-Columbia
423 Lewis Hall
Columbia, MO  65211
kelleyej@health.missouri.edu

References

References marked with * indicate studies included in the meta-analysis.


Copyright © 2015 NCS Pearson, Inc. All rights reserved.


Figure 1. Search for Relevant Articles
### Table 1. Articles Included for Review

<table>
<thead>
<tr>
<th>Citation</th>
<th>Type of Study</th>
<th>Participants</th>
<th>Intervention</th>
<th>Outcomes</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bradshaw, Hoffmann, &amp; Norris (1998)</td>
<td>Alternating treatments single-case design</td>
<td>$n = 2$ Four-year-old boys with language delay.</td>
<td>12 individual sessions in a 6-week period, 30 minutes each. Interactive reading with comparison of two interventions: expansion and cloze versus question, model answer.</td>
<td>Frequency of inferences made by child during session.</td>
<td>Increase in inferences in both conditions, frequency of inferences higher in expansion and cloze condition.</td>
</tr>
<tr>
<td>Desmarais, Nadeau, Trudeau, Filatratu–Veilleux, &amp; Maxés-Fournier (2013)</td>
<td>Multivariate repeated-measures design</td>
<td>$n = 16$ Children with specific language impairment, 4–6-years-old.</td>
<td>10 individual sessions, 1 per week, 15 to 20 minutes each. Interactive reading with scripted questions, prompts, and responses (half referential, half inferential).</td>
<td>PLAI Researcher-created measure of literal and inferential question-answering.</td>
<td>Increase in PLAI scores from pretest to posttest. No increase in literal or inferential question answering for intervention period.</td>
</tr>
<tr>
<td>van Kleeck, Vander Woude, &amp; Hammett (2006)</td>
<td>Experimental prepost</td>
<td>$n = 30$ Children with language impairment, 3–5-years-old.</td>
<td>16 individual sessions in 8 weeks, 15 minutes each. Interactive reading with embedded literal (75%) and inferential (25%) questions and scripted scaffolding and responses.</td>
<td>PLAI, subscores for literal and inferential language. PPVT</td>
<td>Greater gains for treatment group on inferential language, literal language, receptive vocabulary.</td>
</tr>
</tbody>
</table>