Title

Influences of individual, text, and assessment factors on text/discourse comprehension in oral language (listening comprehension)

Journal

Annals of Dyslexia, 71(2)

ISSN

0736-9387

Authors

Kim, Young-Suk Grace
Petscher, Yaacov

Publication Date

2021-07-01

DOI

10.1007/s11881-020-00208-8

Peer reviewed

HHS Public Access

Author manuscript

Ann Dyslexia. Author manuscript; available in PMC 2022 July 01.

Published in final edited form as:

Ann Dyslexia. 2021 July ; 71(2): 218–237. doi:10.1007/s11881-020-00208-8.

Influences of Individual, Text, and Assessment Factors on Text/Discourse Comprehension in Oral Language (Listening Comprehension)

Young-Suk Grace Kim,
University of California at Irvine

Yaacov Petscher
Florida State University

Abstract

We investigated the contributions of multiple strands of factors—individual characteristics (struggling reader status, working memory, vocabulary, grammatical knowledge, knowledge-based inference, theory of mind, comprehension monitoring), a text feature (narrative vs. expository genre), and question types (literal and inferential)—to one's performance on discourse comprehension in oral language (listening comprehension), using data from 529 second graders. Results from explanatory item response models revealed that substantial variance in listening comprehension was attributable to differences between items, texts, and children, respectively. Narrative versus expository genre distinctions explained almost all of the variance attributable to text differences. In contrast, literal versus inferential question distinctions did not explain item responses after accounting for text and reading comprehension status. However, there was a moderation between struggling reader status and question type such that struggling readers had a slightly higher (2%) probability of getting inferential questions right compared to typically developing readers, after accounting for individual and text factors. Struggling readers have a lower probability of accurate item responses than typically developing readers, but the difference disappeared once language and cognitive skills (e.g., working memory, vocabulary) were taken into consideration. The effects of text genre and question type on item responses did not differ as a function of children's language and cognitive skills. Overall these results underscore the importance of considering individual, text, and assessment factors for children's performance in listening comprehension.

Keywords

Assessment; Text; Narrative; Expository; Listening Comprehension; Discourse Comprehension

Text or discourse comprehension, which includes comprehension of both oral texts (listening comprehension) and written texts (reading comprehension), is an essential skill in our modern information-driven society. Unfortunately, however, many children struggle and fail

Corresponding Author, Young-Suk Grace Kim, Professor, School of Education, University of California at Irvine, youngsk7@uci.edu, Phone: (949) 824-2348.

to develop proficient comprehension skills. For example, the National Assessment of Educational Progress (NAEP) in the United States has consistently found that approximately three fourths of students read at or below basic proficiency. In addition, millions of students with learning disabilities also struggle with reading development. Research in the last four decades has made great strides and revealed numerous factors that influence one's discourse comprehension. However, the vast majority of prior work focused on person or individual characteristics (e.g., vocabulary knowledge) without sufficient attention to the roles of other factors such as text features. Furthermore, prior work almost exclusively focused on comprehension of written texts (reading comprehension), not of oral texts (listening comprehension), despite the fact that discourse comprehension includes both (Kintsch, 1988; Van Dijk & Kintsch, 1983). In the present study, we address these gaps in the literature by investigating the contributions of multiple strands of factors—child characteristics (e.g., struggling reader status, working memory), a text feature (i.e., genre: expository vs. narrative), and an assessment feature (i.e., question type: literal vs. inferential)—to one's performance on discourse comprehension in oral language (listening comprehension henceforth), using data from second graders in the United States.

Theoretical Models of Text/Discourse Comprehension and Evidence

Successful text comprehension requires construction of the situation model—a mental representation of the situation described by the text (Graesser, Singer, & Trabasso, 1994; Kintsch, 1988). Constructing an accurate situation model requires highly complex information processing involving construction and integration processes (e.g., see McNamara & Magliano, 2009, for a review). Prior work, both empirical and theoretical, focused on individual/person characteristics that contribute to the complex processes of discourse comprehension (e.g., the simple view of reading [Gough & Tunmer, 1986]; Direct and Inferential Mediation Model of Reading Comprehension [DIME], Cromley & Azevedo, 2007). However, growing evidence suggests the roles of text features and activity/assessment features in discourse comprehension (e.g., Collins et al., 2019; McNamara, Kintsch, Songer, & Kintsch, 1996; Wolfe & Woodwyk, 2010), and these roles have been formally recognized in the direct and indirect effects model of reading (DIER; Kim, 2020). According to DIER, individual characteristics such as working memory, background knowledge, and socio-emotions do influence one's discourse comprehension. In addition, their roles in discourse comprehension are posited to differ as a function of text features (e.g., orthographic and morphological characteristics in written texts, demands on vocabulary, inference, background knowledge). Assessment factors are also posited to play a role in the extent to which one's comprehension is captured. Below is a brief review of literature on each strand—individual characteristics, assessment features (specifically types of questions), and text features.

The Roles of Individual, Assessment, and Text Factors in Discourse Comprehension

Evidence clearly indicates that a number of language and cognitive skills are involved in discourse comprehension processes, including working memory (Daneman & Merikle,

1996; Florit, Roch, & Levorato, 2011; Kim, 2015, 2016; Zwaan & Radvansky, 1998), inhibitory control (Kim & Phillips, 2014), attentional control (Conners, 2009; Kim, 2016), vocabulary (Florit et al., 2011; Kim, 2015, 2016; Strasser & del Rio, 2014), grammatical knowledge (Cain, 2007; Florit et al., 2011, 2014; Kim, 2015, 2016, 2017, 2020; Senechal, Ouellette, & Rodney, 2006), inference-making (inference hereafter; Cain, Oakhill, & Bryant, 2004; Florit et al., 2011; Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Kim, 2016, 2017, 2020; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Tompkins, Guo, & Justice, 2013), perspective taking as measured by theory of mind (Kim, 2015, 2016; Kim & Phillips, 2014), comprehension monitoring (Kim, 2015; Kim & Phillips, 2014; Strasser & del Rio, 2014), and knowledge (topic/content knowledge, McNamara et al., 1996; text structure knowledge, Cain et al., 2004). Not surprisingly, children who struggle with discourse comprehension have lower skills in these language and cognitive domains (Cain & Oakhill, 1999, 2006; Ehrlich, Remond, & Tardieu, 1999; Nation, Clarke, Marshall, & Durand, 2004; Oakhill, 1984).

In regard to assessment and instruction of comprehension, two types of comprehension have been widely distinguished: literal comprehension versus inferential comprehension (Carnine, Silbert, Kame'enui, & Tarver, 2010; Cecil, Baker, & Lozano, 2015; Leslie & Caldwell, 2011; McCormick, 1992; McKenna & Stahl, 2009; Pearson & Dole, 1988; Pearson & Johnson, 1978; Raphael, 1984; Vacca et al., 2009). Literal comprehension refers to one's understanding of what is explicitly stated in the text (Pearson & Johnson, 1978), whereas inferential comprehension is an understanding of what is not explicitly specified but implied in the text—that is, "read[ing] between the lines" (Basaraba, Yovanoff, Alonzo, & Tindal, 2013, p. 354). The literal and inferential taxonomy is widely adopted in assessment of discourse comprehension in normed tasks, high-stakes state-level assessments, and NAEP (e.g., Mazany, Pimentel, Orr, & Crovo, 2015).

Literal comprehension is typically viewed as low-level or shallow comprehension that is necessary for and easier than higher-level inferential comprehension (Alonzo, Basaraba, Tindal, & Carriveau, 2009; Applegate, Quinn, & Applegate, 2002; Carnine et al., 2010; Lapp & Flood, 1983; McCormick, 1992). However, evidence from previous studies is mixed about the difficulty of literal versus inferential comprehension. In some studies, poor comprehenders were found to have a particular difficulty with inferential comprehension questions (Cain & Oakhill, 1999; Davey & Macready, 1985; Holmes, 1987). In another study, Potocki, Ecalle, and Magnan (2013) examined listening comprehension by five-year-old skilled, less skilled, and poor comprehenders. In literal comprehension, performance did not differ between skilled and less-skilled comprehenders, but these groups outperformed poor comprehenders. In inferential comprehension, skilled comprehenders outperformed both less skilled and poor comprehenders. In Miller and Smith's (1985) study, children in Grades 2 to 5 showed no differences in performance levels as a function of the type of comprehension questions.

Finally, text features also influence discourse comprehension. Texts vary in the demands of language and cognitive skills—some texts include advanced vocabulary and/or sentence structures and/or require a greater extent of inference, perspective taking, or content knowledge. These features tend to covary by genre (e.g., narrative vs. expository). Narrative

texts have received much attention with a long history in various fields, but expository or informational texts have garnered their due attention relatively recently, particularly for developing children (e.g., Duke, 2000). Scholars have documented differences in language characteristics and structural aspects between narrative and informational texts (Derewianka, 1990; Duke & Kays, 1998; Goldman & Rakestraw, 2000; Stein & Trabasso, 1981). Successful comprehension of narrative texts tends to involve processes related to achieving coherence in thematic and causal structure that typically happen though time, whereas comprehension of expository texts tends to involve creating a coherent representation of the text content, including causal structure, and integration of text content with relevant content knowledge (Brewer, 1980; Graesser, León, & Otero, 2002; Graesser et al., 1994; Trabasso & Magliano, 1996). Comprehension of narrative texts was found to be easier than expository texts for children (Best, Floyd, & McNamara, 2008; Williams, Hall, & Lauer, 2004) and adults (Wolfe & Woodwyk, 2010), most likely due to greater exposure or familiarity (Duke, 2000), less varied text structure, and less demand on knowledge on a given topic (e.g., McKeown, Beck, Sinatra, & Loxterman, 1992; McNamara et al., 1996; Wolfe & Mienko, 2007).

Gaps in the Literature and the Present Study

Prior work reviewed above has provided rich insight into the complexity of discourse comprehension. However, there are several noteworthy gaps. First, discourse comprehension in the vast majority of prior work was conducted on reading comprehension although work on listening comprehension has been growing in recent years. Listening comprehension is a necessary precursor and foundation for reading comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990), and therefore, listening comprehension merits attention. Theoretically, discourse comprehension does not differentiate reading versus listening comprehension in terms of processes (e.g., Kintsch, 1988; McNamara & Magliano, 2009) with an exception of word reading processes involved in reading comprehension (e.g., Gough & Tunmer, 1986; Kim, 2020), and recent evidence revealed that highly similar language and cognitive skills contribute to reading comprehension and listening comprehension (e.g., for listening comprehension, see Florit et al., 2014; Kim, 2015, 2016, 2020; Lepola et al., 2012; Strasser & del Rio, 2014; Tompkins et al., 2013; for reading comprehension, see, e.g., Cain et al., 2004; Oakhill et al., 2003, 2005; Savage, Cornish, Manly, & Hollis, 2006).

Another important gap is an understanding of how the abovementioned multiple strands of factors *together* influence discourse comprehension. The roles of individual, text, and assessment factors in comprehension have largely been studied in disparate lines of work. Consequently, we have limited knowledge of how these factors contribute to discourse comprehension in the context of one another and whether their contributions vary as a function of each other (i.e., moderation). For example, one's inferencing ability (individual factor) may have a larger effect on inferential comprehension questions compared to literal comprehension questions (e.g., Eason et al., 2012). A recent study on reading comprehension with students in Grade 4 showed an interaction between an assessment factor (open-ended versus multiple choice response format) and an individual factor (language knowledge) such that language knowledge had a greater effect on open-ended items than

multiple choice items in reading comprehension (Collins et al., 2020). Furthermore, text genre did not explain item accuracy in reading comprehension after accounting for language and cognitive skills, and item response format (Collins et al., 2020). To our knowledge, no previous studies have investigated the roles of individual, text (genre), and assessment (question types) factors simultaneously in listening comprehension.

To address these gaps in the literature and to develop a deeper understanding of the roles of multiple strands of factors in discourse comprehension, we investigated how individual characteristics (reading comprehension status, and language and cognitive skills), a text feature (genre: expository vs. narrative), and question types (literal vs. inferential comprehension) relate to one's performance on listening comprehension, using data from developing readers in Grade 2. With respect to the individual factors, children's reading comprehension status—poor or typical, with "poor" operationalized as a standard score of 85 or below in a reading comprehension task—was included for three reasons. The first reason was to confirm its relation with listening comprehension. Given the established relation between reading comprehension and listening comprehension (Catts, Adlof, & Ellis Weismer, 2006; Florit & Cain, 2011; Hoover & Gough, 1990; Kim, 2015, 2017), many poor reading comprehenders would likely be poor listening comprehenders as well. Moreover, if reading comprehension and listening comprehension draw on essentially the same language and cognitive skills except for those involved in word reading processes (e.g., phonological, orthographic, and semantic aspects), then once language and cognitive skills are accounted for, there would not be any difference in the probability of answering listening comprehension items correctly between struggling and typically developing reading comprehenders. The second reason was to explore potential moderations of reading comprehension status with a text feature and question types—that is, whether struggling reading comprehenders have greater difficulty with expository texts than with narrative texts, and with inferential comprehension questions than with literal comprehension questions. The final reason was to reflect practices in schools where decisions about instruction (e.g., grouping of students for differentiated instruction) and/or referral are typically made based on students' performance on reading comprehension, not listening comprehension.

The following four research questions guided the present study: (1) How much variance in listening comprehension is attributable to differences between individuals, texts (passages), and items?; (2) Do reading comprehension status (poor vs. typical comprehender), text genre (expository vs. narrative passage), and assessment question type (literal vs. inferential item) relate to listening comprehension?; (3) Do the effects of text genre and question type on listening comprehension vary for poor reading comprehenders versus typical reading comprehenders? Does the effect of text genre on listening comprehension vary by question type?; and (4) Do text genre, question type, and reading comprehension status explain children's performance on listening comprehension after accounting for children's language and cognitive skills (working memory, vocabulary, grammatical knowledge, inference, perspective taking, and comprehension monitoring)? Do the effects of text genre and question type on listening comprehension vary by children's language and cognitive skills?

Note that the language and cognitive skills included in this study was informed by prior evidence, and considering the practical constraint of working in the schools where a very

large assessment battery is often not feasible. We anticipated that listening comprehension of expository texts would be more difficult than narrative texts for children, in line with previous work in reading comprehension, and that children who struggle with reading comprehension (i.e., poor reading comprehenders) would have lower performance than typical reading comprehenders on listening comprehension. However, we did not have a clear hypothesis about the difficulty level of literal comprehension compared to inferential comprehension items, given mixed findings in prior work (e.g., Cain & Oakill, 1999 versus Miller & Smith, 1985). We expected that once language and cognitive skills (e.g., working memory, vocabulary, inference) were accounted for, reading comprehension status would no longer explain children's performance on listening comprehension. Finally, we did not have clear a priori hypotheses about whether other language and cognitive skills would moderate the effects of text genre and question type.

Method

Participants

Participants included 529 second graders (53% males) in a Southeastern state in the United States. These children were composed of three cohorts of students (*N*s = 165, 185, and 179 in each cohort) from the same schools, who were assessed in three consecutive academic years in an identical manner. Data from the first two cohorts were used in an article that examined structural relations of language and cognitive skills (Kim, 2017). Seventy-two percent of the sample children qualified for free or reduced lunch, a proxy variable for poverty status. The racial/ethnic breakdown was as follows: 53% White, 34% Black, 5% Hispanic, .9% Asian/Pacific Islander, and 5% identified as two or more races/ethnicities. Only 1% of the children ($n = 7$) were classified as English language learners. Approximately 13% of the children received speech services and 1% were identified to have language impairment. All children were included in the analysis.

Measures

Unless otherwise noted, children's responses to items in all tasks were scored dichotomously.

Reading Comprehension.—Children's reading comprehension was measured by two normed tasks: the Reading Comprehension subtest of the Wechsler Individual Achievement Test-III (WIAT-III; Wechsler, 2009) and the Passage Comprehension subtest of the Woodcock Johnson-III (WJ-III; Woodcock, McGrew, & Mather, 2001). In the WIAT-III, the child was asked to read narrative and expository passages and answer multiple choice comprehension questions. In the WJ-III task, the child was asked to read sentences and passages and fill in blanks. Cronbach's alpha estimates were .82 and .83 for the WIAT-III and WJ-III, respectively. Children with poor reading comprehension were identified as those whose standard score was 85 or below in either of the reading comprehension tasks.

Listening comprehension.—Children's listening comprehension of narrative and expository texts was assessed by the Narrative Comprehension subtest of the Test of Narrative Language (TNL; Gillam & Pearson, 2004) and an experimental expository task,

respectively. In the TNL Narrative Comprehension subtest (Tasks 1, 3, & 5 are comprehension tasks), the child heard three narrative stories and was asked open-ended comprehension questions for each story (a total of 30 questions). Following the TNL manual, the majority of items were scored using a dichotomous scale of 0 or 1, but some items were scored using a trichotomous scale of 0, 1, or 2 for a total possible maximum score of 40. Cronbach's alpha was .80.

The experimental expository comprehension task was composed of three expository passages from the Qualitative Reading Inventory-5 (QRI-5; Leslie & Caldwell, 2011). Titles of the passages were as follows: *Changing Matter* (140 words), *Whales and Fish* (200 words), and *Where do People Live?* (282 words). After listening to each passage, the child was asked comprehension questions (a total of 24 questions across the three passages). Cronbach's alpha was .76.

Comprehension questions in the TNL and experimental expository task were classified into literal and inferential questions. Literal questions required children to recall explicitly stated information from the text (e.g., "What was the girl's name?" in the Test of Narrative Language Task 1; "What three things does all matter have?" in the *Changing Matter* expository text), whereas inferential questions required inferring information that was not explicitly stated in the text (e.g., "What was the problem in the story?" in the Test of Narrative Language Task 1; "What do you think causes matter to change form?" in the *Changing Matter* expository text).

To distinguish literal from inferential question types in the experimental expository task, we used the designation of question types identified by the authors of the QRI-5 (Leslie & Caldwell, 2011). For the TNL, literal and inferential question types were not identified by the TNL authors, and thus, the questions were coded by the present study's research team into literal and inferential questions. If the correct response was explicitly stated in the text, the question was coded as a literal question; otherwise, the question was coded as an inferential question. Agreement rate was 100% between two coders (first author and a graduate student). There were a total of 39 literal questions (25 in the TNL) and 15 inferential questions (5 in the TNL) across the TNL and the QRI-5 passages.

Knowledge-based inference.—Knowledge-based inferencing skill was measured by the Inference task of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999). In this task, the child heard 2- to 3-sentence stories and was asked a question that required inference using background knowledge (e.g., *Mother called to four-year-old Sandra and says 'Be sure to bring your bathing suit. And don't forget your shovel and bucket.' Where are they going?*). Two practice items and 25 test items were included. Test administration discontinued after five consecutive incorrect items, following the assessment protocol. Cronbach's alpha was .89.

Perspective taking (theory of mind).—A theory of mind task, false belief task specifically, was used. Studies have shown that first-order theory of mind develops around age 4 (Wellman, Cross, & Watson, 2001) while second-order theory of mind develops around age 5 to 7 (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994).

Considering the developmental phase of children in our sample, three second-order scenarios were used. Second-order scenarios examine a child's ability to infer a story character's mistaken belief about another character's knowledge (e.g., John may think, "Aaron believes that Jane knows that there is a bake sale"; see Arslan, Hohenberger, & Verbrugge, 2017) and, therefore, tap one's complex reasoning skill, particularly related to perspectives. The three scenarios involved the context of a bake sale, visit to a farm, and going out for a birthday celebration. These scenarios were presented with a series of illustrations, followed by questions. There were six questions per scenario for a total of 18 questions. Cronbach's alpha was .79.

Comprehension monitoring.—An inconsistency detection task was used (e.g., Baker, 1984; Cain et al., 2004; Kim & Phillips, 2014). The child heard a short scenario and was asked to identify whether the story made sense or not (e.g., *Susan's favorite color is green. Her bag is green. Her pants are green. Susan's favorite color is red.*). If the child indicated that the story did not make sense, she was asked to provide a brief explanation and to fix the story so that it made sense. There were two practice items and nine experimental items. Consistent (three items) and inconsistent stories (six items) were randomly ordered. For all nine items, accuracy of the child's answer about whether a scenario was consistent or inconsistent was dichotomously scored. For the six inconsistent stories, the accuracy of the child's explanation and repair of the story were also dichotomously scored for each item. If the child correctly responded to an inconsistent story, the total maximum possible score for the item was 3—one for correctly identifying inconsistency, one for providing a correct explanation, and one for an accurate repair; thus, the total possible score was 21. Note that the correlation of the score accounting for the repair versus not was extremely high, and therefore, the score accounting for the repair was used in the study (see Kim, 2017). Cronbach's alpha was .77.

Vocabulary.—An expressive measure, the Picture Vocabulary subtest of the WJ-III (Woodcock et al., 2001) was used. In this task, the child was asked to identify pictured objects. Test administration discontinued after six consecutive incorrect items. Cronbach's alpha was .77.

Grammatical knowledge.—The Grammaticality Judgement task of CASL (Carrow-Woolfolk, 1999) was used. The child heard a sentence (e.g., *The children is running*) and was asked whether the sentence was grammatically correct. If incorrect, the child was asked to correct the sentence. Test administration discontinued after five consecutive incorrect items. Cronbach's alpha was .90.

Working memory.—A listening span task (Daneman & Merikle, 1996; Kim, 2015, 2016) was used to measure working memory. In this task, the child was presented with a short sentence involving common knowledge familiar to children and was asked to identify whether the heard sentence (e.g., Apples are blue) was correct or not. After hearing multiple sentences (i.e., two to four), the child was asked to identify the last word of each sentence. There were four practice items and 13 experimental items. Children's yes/no responses regarding the veracity of the statement were not scored, but their responses on the last words

in correct order were given a score of 0 to 2: correct last words in correct order were given 2 points, correct last words in incorrect order were given 1 point, and incorrect last words were given 0 point. The total possible score was 26. Testing discontinued after three incorrect responses. Cronbach's alpha was estimated to be .74.

Procedures

Children were individually assessed in a quiet space in the school. The assessment battery was administered in several sessions with each session 30 to 40 minutes long.

Data Analysis

A series of explanatory item response models (EIRMs; De Boeck & Wilson, 2004) were used to understand the extent to which variation in listening comprehension item accuracy was due to child-level differences compared to item-level differences, passage-level differences, classroom-level differences, and school-level differences. EIRMs are a form of generalized linear mixed models that blend multilevel and psychometric traditions to evaluate person-level ability and item-level accuracy along with predictors of individual differences on both sides (Petscher, Compton, Steacy, Kinnon, & McNeish, 2019). Although a variety of EIRMs appear in the literature, our approach used a variation of the double-explanatory model. The double-explanatory model refers to an EIRM that uses item and person covariates to explain variance in item accuracy at the item and person levels. Our statistical modeling included five total random effects (i.e., child, item, passage, classroom, and school), and covariates were included for the child, item, and passage effects. Classroom and school effects were included to account for the shared environment and nesting structure of the data; however, their inclusion was primarily so that the standard errors for the fixed effects were appropriately estimated. No specific classroom or school predictors were included. It is important to note that because EIRMs are a form of psychometric modeling (i.e., item response theory models), the intercept can be interpreted to understand person-level ability and item-level difficulty. The focus of our research questions primarily lies with understanding person-level ability and explanations of variance; thus, our interpretations in the unconditional model will focus on person-level log-odds of success and not the item-level difficulty.

A total of five EIRMs were estimated to address the research questions. To address the first research question, an unconditional model estimated the mean log-odds of item-level accuracy as well as the variance for each of the five levels in the model (i.e., item-level differences, passage-level differences, child-level differences, classroom-level differences, and school-level differences). Intraclass correlations were computed from the random effects to understand what percentage of the variance was due to each of the five levels.

Research questions two to four were addressed by fitting four conditional EIRMs. The second research question was examined in Model 1, which included the dichotomous variables to explain passage effects (i.e., expository vs. narrative) and item effects (i.e., literal vs. inferential questions) as well as the indicator of whether children had poor reading comprehension. The third research question was addressed by Model 2, which was built on Model 1 by including two-way and three-way interactions among the expository versus

narrative passage variable, literal versus inferential question variable, and poor versus typical comprehender variable. The fourth research question was addressed by Models 3 and 4. Model 3 added to Model 2 with grand-mean centered child-level indicators of inferencing, theory of mind, comprehension monitoring, vocabulary, grammatical knowledge, and working memory. Model 4 included the child-, item-, and passage-level main effects from Model 3 along with two-way and three-way interactions of the item and passage variables with child-level language and cognitive skills. Each of Models 1–4 was compared to the unconditional model via pseudo r-squared statistics to understand the respective proportion of variance explained at each level based on the included covariates. All analyses were conducted using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014).

Results

Preliminary Analysis and Descriptive Statistics

A preliminary review of the data showed that less than 1% of the responses were missing for any of the measured variables. Little's missing completely at random (MCAR) test suggested that all missing data met reasonable assumptions for MCAR, χ^2 (21) = 19.52, p = .552; thus, using full information maximum likelihood for model estimation was appropriate. Descriptive statistics and correlations are reported in Table 1 and showed that 27% of children in this sample met the criteria for poor reading comprehension. The sample children's standard scores in normed tasks, vocabulary (M = 96.78, SD = 10.29), knowledge-based inference (M = 92.70, SD = 12.96), grammatical knowledge (M = 95.50, SD = 13.16), and TNL comprehension (M = 8.47, SD = 2.86 where norm mean is 10), were all within the normal range of development. Correlations among the reading and language measures were weak to moderate in strength, ranging from r = .21 between comprehension monitoring and working memory to r = .67 between knowledge-based inference and the narrative comprehension (TNL). All measures were negatively correlated with the dichotomous indicator of poor reading comprehension ($-.52 < r < -.22$) meaning that a standard score ≤ 85 on either or both of the reading comprehension tasks was associated with lower performance on language and cognitive measures.

Research Question 1: How much variance in listening comprehension is attributable to differences between individuals, passages or texts, and items?

Results from the unconditional model showed that the mean log-odds of item-level accuracy was 0.41 (p = .46). This estimated value indicated that the chance of correctly responding to a given item across the six passages (three narrative and three expository passages) was, on average, .60, which was close to the observed mean percentage correct of 58.7% in the sample data. Random effects for the unconditional model are reported in Table 2, which shows that for the five random effects specified, 33% of the variance in item responses was due to between-item differences followed by 18% due to between-passage differences, 8% due to between-person differences, 2% due to between-classroom differences, and 0% due to between-school differences. The logit-scale variance, a fixed quantity, represented 39% of the total variance. Because the school-level variance was estimated as 0.00, it was removed as a random effect from the conditional models.

Research Question 2: Do reading comprehension status (poor vs. typical reading comprehender), text genre (narrative vs. expository), and question type (literal vs. inferential) relate to listening comprehension?

As shown in conditional Model 1 in Table 3, passage type (narrative vs. expository), question type (literal vs. inferential), and reading comprehension status (poor vs. typical) were significantly related to the log-odds of listening comprehension accuracy. The negative effect for expository passages (−2.37, $p < .001$) indicated that expository passages were harder for children than narrative passages. The intercept value of 1.83 is partially referent to narrative passages, which, when converted to a probability, reflects that children had a .86 chance of correctly answering a question from narrative text. The fitted log-odds value for expository passages was −0.54 (i.e., 1.83 + −2.37) and translated to a .37 probability of success on items from expository text. Poor reading comprehension was negatively related to listening comprehension performance such that poor comprehenders' fitted log-odds of .98 (i.e., 1.83 + −.85) equated to a .73 probability of a correct response compared to .86 for typical comprehenders. The inclusion of the three variables in Model 1 resulted in 10% of the child variance explained, 93% of the classroom variance explained, 1% of the item variance explained, and 98% of the passage variance explained (Table 2). We further explored the nature of the large pseudo R^2 for passage variance by estimating item level difficulty differences between the passage types. It was plausible that text genre explaining between-passage differences was less to do with text complexity or features of genre itself and more to do with the respective difficulty of items for narrative and expository passages. The average percentage of items correct for narrative passages was 74.4% ($SD = 43.6\%$) compared to 39.0% ($SD = 48.8\%$) for expository passages. The standardized effect size difference between these two aggregate difficulty values was $d = 0.76$, a large practically important effect in item difficulties.

Research Question 3: Do the effects of text genre and question type on listening comprehension vary for poor reading comprehenders versus typical reading comprehenders? Does the effect of text genre on listening comprehension vary by question type?

Conditional Model 2 (Table 3) included interactions among the three main effects (expository/narrative passage type, literal/inferential question type, and poor/typical reading comprehension status). Results showed a significant effect for the interaction between question type and reader status (−0.29, $p = .04$; Table 3). The direction of the coefficients for the interaction and main effects pointed to poor comprehenders having a lower probability of a correct response on literal questions compared to inferential questions. However, we refrain from further interpretation of this interaction until the Model 3 results (see below), a more comprehensive model that includes children's language and cognitive skills. The inclusion of interactions did not appreciably change the estimated variance components or pseudo r-squared statistics compared to Model 1.

Model 3 included language and cognitive skills, and results (Table 3) showed statistically significant effects for knowledge-based inference (0.02, $p < .001$), theory of mind (0.07, $p < .001$), comprehension monitoring (0.05, $p < .001$), vocabulary (0.01, $p = .001$), grammatical knowledge (0.01, $p = .003$), and working memory (0.02, $p = .004$). It is of note

that once these language and cognitive skills were included in the model, the effect of poor reading comprehension was no longer statistically significant (0.18, $p = .23$). The inclusion of language and cognitive predictors resulted in 65% of the child-level variance explained.

The interaction between comprehension question type and poor reading comprehension status remained statistically significant (−0.29, $p = .04$) after accounting for language and cognitive skills. As shown in Figure 1, poor reading comprehenders had a .82 probability of a correct response to literal items compared to .92 for inferential items, after controlling for the language and cognitive skills. Typical reading comprehenders had a .83 probability of a correct response on literal items compared to .90 for inferential questions.

Research Question 4: Do text genre, question type, and reading comprehension status explain children's performance on listening comprehension after accounting for children's language and cognitive skills? Do the effects of text genre and question type on listening comprehension vary by children's language and cognitive skills?

Model 4 included interaction terms of language and cognitive skills with genre (expository vs. narrative) and question type (literal vs. inferential). However, results yielded no statistically significant interactions ($ps \geq .08$; Table 3).

Discussion

The primary goal of this study was to investigate the relations of multiple strands of factors —child, text, and assessment (question types)—to listening comprehension, using data from second graders. Overall results corroborated the hypothesized roles of these factors, but nuances were revealed.

One of the striking findings in the present study is a large amount of variance attributable to between-item differences (33%). A similar result was also reported in a study on reading comprehension (Collins et al., 2020). Together, these results indicate that differences between items play a large role in one's performance on discourse comprehension tasks – that is, children's performance on comprehension tasks can vary depending on items to some extent. In the present study, we included literal and inferential questions as an item (or assessment) feature, and found that after accounting for genre and reading comprehension status, there was no difference between literal versus inferential comprehension questions. This finding is discrepant with previous studies of reading comprehension, which reported that literal comprehension questions are easier than inferential questions (Alonzo et al., 2009; Basaraba et al., 2013). Note though that our findings are on listening comprehension, not reading comprehension, and our results were after accounting for poor reader status and a text feature, genre, whereas previous studies did not account for these factors. Furthermore, there was a moderation effect such that poor reading comprehenders had a slightly higher probability than typical comprehenders of getting inferential comprehension questions correct, once language and cognitive skills were accounted for. This result appears discrepant from previous studies that found differences in both inferential and literal types of questions (Potocki et al., 2013) or that reported no differences based on question type (Miller & Smith, 1985). However, results of prior work and the present study cannot be directly compared because of an important difference—in the present study, children's

language and cognitive skills (working memory, vocabulary, grammatical knowledge, knowledge-based inference, perspective taking, and comprehension monitoring) were controlled whereas in previous studies, they were not. In addition, although the interaction was statistically significant, the effect was very small (2% probability difference), and the statistical significance was likely due to a large sample size in the present study. Thus, the present results indicate that a child's score in listening comprehension was not likely to be considerably different even if he or she was given exclusively literal or inferential comprehension questions. However, this certainly does not imply that measuring both literal and inferential comprehension questions is unimportant. Instead, the results suggest that a large amount of variance in listening comprehension is attributable to between-item differences, and only a small amount of variance (4%) is explained by literal and inferential question types.

These results indicate a need for being cognizant of the roles of item differences in children's performance on comprehension tasks and a need for careful attention to items in comprehension tasks. Although items might appear similar in terms of demands (e.g., requiring recall of texts), a careful look (e.g., in a causal network analysis) might reveal differential roles of the requested information in the texts, which might influence student performance. Results also indicate that measuring comprehension skill with multiple tasks and varying item types, and using a latent variable as an analytical approach would help capture listening comprehension with precision. Finally, these results warrant future work to shed light on assessment of discourse comprehension. For example, future work is warranted to expand our understanding of various item features in reading and listening comprehension. Future work should also investigate an assessment's response format (open-ended vs. multiple choice) as it has been shown to play a role in reading comprehension performance (Collins et al., 2020; Reardon, Kalogrides, Fahle, Podolsky, & Zárate, 2018). Response format might also be a factor in listening comprehension performance, but we could not examine this because all the items in the present study had an open-ended response format.

The present study also revealed that a substantial amount of variance is attributable to between-passage differences (18%), and it is marked that text genre (narrative vs. expository) explained almost all the variance due to between-passage differences (98% of 18%), indicating the important role of text genre in discourse comprehension (Alvermann, Hynd, & Qian, 1995; Wolfe, 2005). The explanatory role of text genre indicates that genre captures many differences in texts at least for children in Grade 2. However, it is difficult to completely separate genre from other features (e.g., content, item difficulty, text complexity). One possibility of the substantial role of genre is that item difficulties or text complexities systematically differed along the lines of the genre. As shown in the Results section, students' performance on narrative texts was lower than on informational texts, indicating differences in item difficulties by genre. Text complexity as measured by Lexile ranged from 210 to 600 for the informational texts and from 410 to 800 for the narrative texts, but two out of three passages in narrative genre had identical values with two informational passages. That is, only one passage in narrative genre had higher Lexile, and one informational passage had lower Lexile. It appears that text complexity measured by Lexile is not likely the primary explanation for the observed role of text genre. Future

studies with an experimental design that manipulates factors (e.g., see Wolfe & Mienko, 2007 for a study with adults) are needed to further elucidate and tease out the roles of specific text features on one's performance on comprehension tasks.

The greater difficulty of expository texts (.37 probability of success on expository text items) than narrative texts (.86 probability of success on narrative text items) is convergent with some previous studies in reading comprehension (Best et al., 2008; Williams et al., 2004; Wolfe & Woodwyk, 2010), but not with Collins et al.'s study (2020), which revealed no significant differences in performance as a function of text genre. The difference of listening comprehension in the present study versus reading comprehension in previous studies does not explain the discrepancy given the mixed findings in reading comprehension studies. Our results add to the literature by showing a difference in difficulty between narrative and expository texts after controlling for the type of comprehension questions and children's language and cognitive skills. Prior work suggested that the difference in difficulty between narrative texts and expository texts is likely due to multiple factors—differences in children's familiarity (Duke, 2000), density of information and language demands (Meyer & Ray, 2011), reliance on topic knowledge (Wolfe & Woodwyk, 2010), and variability in text structures (Duke, 2000; Williams et al., 2004). Overall, these results indicate differences in test scores in listening comprehension when the same children are given narrative versus expository texts and, therefore, underscore the importance of including both narrative and expository texts when measuring children's listening comprehension.

A surprising finding in the present study was a smaller-than-expected amount of variance attributable to between-child differences (8%). To our knowledge, no previous studies have examined variances attributable to child, text, and assessment factors in children's listening comprehension; therefore, the present result cannot be compared. However, a previous study on reading comprehension has shown that a substantial amount of variance in children's item responses was attributable to between-child differences (Collins et al., 2020). It is unclear what explains the difference in the amount of variance attributable to between-child differences in listening comprehension versus reading comprehension. One apparent difference in listening versus reading comprehension is availability of texts—in reading contexts, texts are available as long as access to texts is allowed after the child finishes reading in the assessment protocol, whereas in oral contexts, texts are not available. However, whether, and if so how, this fact explains the difference in the amount of variance attributable to child factors remains an open question.

The smaller amount of variance attributable to between-child differences compared to item features and text (genre) features should not be taken to indicate that child characteristics are less important to discourse comprehension than passage and item factors. Convergent with previous studies (e.g., Author, 2015, 2020; Cain & Oakhill, 1999; Florit et al., 2014), children's reading comprehension status and their language and cognitive skills (working memory, vocabulary, grammatical knowledge, knowledge-based inference, perspective taking, and comprehension monitoring) were related to accuracy in children's responses. In fact, the included language and cognitive skills explained 65% of variance attributable to between-person differences. As expected, children's language and cognitive skills were related to their performance on listening comprehension. Also as expected, struggling

readers had a lower likelihood of getting items correct in listening comprehension. Furthermore, the difference between poor readers and typically developing readers disappeared once language and cognitive skills were taken into consideration, indicating that the included language and cognitive skills explain the difference between poor readers and typically developing readers in listening comprehension. Overall, these results, in conjunction with prior work, indicate that child factors do influence one's performance in listening comprehension, but text features, such as genre, and item features should be taken into consideration for their roles in discourse comprehension.

Beyond the main effects of children's language and cognitive skills, we also investigated their potential moderation with question type and genre. One might speculate that, for example, inference may have a greater effect on inferential comprehension questions compared to literal comprehension questions (Eason et al., 2012). Similarly, theory of mind, one's understanding of others' mental states such as thoughts, intentions, and emotions, may be more relevant to narrative text comprehension because narrative texts involve interactions among characters, and thus, theory of mind may be critical to successful narrative comprehension. However, none of the interactions were statistically significant, suggesting that the various language and cognitive skills did not differentially influence, based on genre or question type, the accuracy of a child's response to listening comprehension questions.

Limitations, Implications, and Future Directions

The present study revealed the roles of multiple strands of factors—individual, text, and question type factors—in children's listening comprehension performance. However, limited features were examined, particularly for the assessment strand as literal vs. inferential question type explained only a small amount of the variance attributable to between-item differences. In addition, although we included relatively comprehensive language and cognitive skills, additional variables merit attention. In particular, future work should include topic knowledge for its potential moderation with genre, given prior evidence on a greater role of topic knowledge in expository texts (Kaefer, Neuman, & Pinkham, 2015; McKeown et al., 1992).

The null effect of poor reading comprehender status once language and cognitive skills were controlled for supports that the poor comprehender status is largely explained by these language and cognitive skills. However, poor reading comprehenders versus typical reading comprehenders status is influenced by not only by comprehension skills, but also by decoding/word reading skills. Therefore, the poor comprehender group is composed of children with heterogeneous profiles of skills, including their primary weaknesses in word reading, language comprehension, or both.

Results of the present study suggest a need for a deeper understanding of factors that influence children's discourse comprehension, listening comprehension in particular. The complexity of assessment of discourse comprehension has been widely recognized in reading comprehension (Collins et al., 2020; Francis, Fletcher, Catts, & Tomblin, 2005; Francis et al., 2006), but little attention has been paid to its implications for listening comprehension. Specifically, the large amount of variance attributable to between-item

differences indicates a need to better understand item features that influence one's performance in listening comprehension tasks. In practice, the effects of narrative versus expository texts indicate a need for including both types of texts in measurement of listening comprehension. In other words, to accurately measure one's listening comprehension, ideally multiple tasks that include both narrative and expository texts with a variety of item features are needed because relying on a single task or format is likely to paint only a partial picture of one's listening comprehension skill. The use of multiple tasks, however, is not often feasible in many settings including schools due to limited time and resources, and therefore efforts are needed to develop accurate but efficient measurement of discourse comprehension that is aligned with theoretical models.

Acknowledgements

This research was supported by the grant from the Institute of Education Sciences, US Department of Education (R305A130131 & R305A180055) and National Institute of Child Health and Human Development (P50HD052120). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency. The author(s) wish to thank participating schools and children.

References

Alonzo J, Basaraba D, Tindal G, & Carriveau RS (2009). They read, but how well do they understand? An empirical look at the nuances of measuring reading comprehension. Assessment for Effective Intervention, 35(1), 34–44. doi:10.1177/1534508408330082

Alvermann DE, Hynd CE, & Qian G (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. Journal of Educational Research, 88, 146–154.

Applegate MD, Quinn KB, & Applegate AJ (2002). Levels of thinking required by comprehension questions in informal inventories. The Reading Teacher, 56, 174–180.

Arslan B, Hohenberger A, & Verbrugge R (2017). Syntactic recursion facilitates and working memory predicts recursive theory of mind. PloS ONE, 12(1), doi:10.1371/journal.pone.0169510

Baker L (1984). Children's effective use of multiple standards for evaluating their comprehension. Journal of Educational Psychology, 76, 588–597.

Basaraba D, Yovanoff P, Alonzo J, & Tindal G (2013). Examining the structure of reading comprehension: do literal, inferential, and evaluative comprehension truly exist? Reading and Writing: An Interdisciplinary Journal, 26, 349–379. doi:10.1007/s11145-012-9372-9

Bates D, Mächler M, Bolker B, & Walker S (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Best RM, Floyd RG, & McNamara DS (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. Reading Psychology, 29(2), 137–164. doi:10.1080/02702710801963951

Brewer WF (1980). Literary theory, rhetoric, and stylistics: Implications for psychology. In Shapiro RJ, Bruce BC, & Brewer WF (Eds.), Theoretical issues in reading comprehension (pp. 221–239). Hillsdale, NJ: Erlbaum.

Cain K (2007). Grammatical awareness and reading ability: Is there any evidence for a special relationship? Applied Psycholinguistics, 28, 679–694. doi:10.1017/S0142716407070361

Cain K, & Oakhill JV (1999). Inference making ability and its relation to comprehension failure in young children. Reading and Writing: An Interdisciplinary Journal, 11, 489–503.

Cain K, & Oakhill JV (2006). Profiles of children with specific reading comprehension difficulties. British Journal of Educational Psychology, 76, 683–696. doi:10.1348/000709905X67610

Cain K, Oakhill J, & Bryant P (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. Journal of Educational Psychology, 96, 31–42. doi:10.1037/0022-0663.96.1.31

Carnine DW, Silbert J, Kame'enui EJ, & Tarver SG (2010). Direct instruction reading (5th Ed.). Boston, MA: Merrill.

Carrow-Woolfolk E (1999). Comprehensive assessment of spoken language. Bloomington, MN: Pearson Assessment.

Catts H, Adlof SM, & Ellis Weismer SE (2006). Language deficits in poor comprehenders: A case for the simple view of reading. Journal of Speech, Language, and Hearing Research, 49(2), 278–293. doi:1092-4388/06/4902-0278

Cecil NL, Baker S, & Lozano AS (2015). Striking a balance: A comprehensive approach to early literacy (5th Edition). Scottsdale, AZ: Holcomb Hathaway.

Collins AA, Compton DL, Lindstrom ER, & Gilbert JK (2020). Performance variations across reading comprehension assessments: Examining the unique contributions of text, activity, and reader. Reading and Writing: An Interdisciplinary Journal. doi:10.1007/s11145-019-09972-5

Conners FA (2009). Attentional control and the Simple View of reading. Reading and Writing: An Interdisciplinary Journal, 22, 591–613. doi:10.1007/s11145-008-9126-x

Cromley J, & Azevedo R (2007). Testing and refining the direct and inferential mediation model of reading comprehension. Journal of Educational Psychology, 99, 311–325. doi:10.1037/0022-0663.99.2.311

Daneman M, & Merikle PM (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin & Review, 3, 422–433. [PubMed: 24213976]

Davey B, & Macready GB (1985). Prerequisite relations among inference tasks for good and poor readers. Journal of Educational Psychology, 77, 539–552.

De Boeck P, & Wilson M (2004). A framework for item response models. In De Boeck P & Wilson M (Eds.), Explanatory item response models: A generalized linear and nonlinear approach (pp. 3–41). New York, NY: Springer.

Derewianka B (1990). Exploring how texts work. Rozelle, NSW, Australia: Primary English Teaching Association.

Duke NK (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. Reading Research Quarterly, 35, 202–224.

Duke NK, & Kays J (1998). "Can I say 'once upon a time'?" Kindergarten children developing knowledge of information book language. Early Childhood Research Quarterly, 13, 295–318.

Eason S, Goldberg LF, Young KM, Geist MC, & Cutting LE (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. Journal of Educational Psychology, 104, 515–528. [PubMed: 26566295]

Ehrlich MF, Remond M, & Tardieu H (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. Reading and Writing, 11, 29–63.

Florit E, & Cain K (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? Educational Psychology Review, 23, 553–576. doi:10.1007/s10648-011-9175-6

Florit E, Roch M, & Levorato MC (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. Discourse Processes, 48, 119–138. doi:10.1080/0163853X.2010.494244

Florit E, Roch M, & Levorato MC (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. Reading and Writing: An Interdisciplinary Journal, 27, 793–817. doi:10.1007/s11145-013-9464-1

Francis DJ, Fletcher JM, Catts H, & Tomblin JB (2005). Dimensions affecting the assessment of reading comprehension. In Paris SG & Stahl SA (Eds.), Children's reading comprehension and assessment (pp. 35–49). Mahwah, NJ: Erlbaum.

Francis DJ, Snow CE, August D, Carlson CD, Miller J, & Iglesias A (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. Scientific Studies of Reading, 10, 301–322. doi:10.1207/s1532799xssr1003_6

Gillam RB, & Pearson NA (2004). Test of Narrative Language. Austin, TX: Pro-Ed.

Goldman SR, & Rakestraw JA (2000). Structural aspects of constructing meaning from text. In Kamil ML, Mosenthal PB, Pearson PD, & Barr R (Eds.), Handbook of reading research (Vol. 3, pp. 311–335). Mahwah, NJ: Erlbaum.

Gough PB, & Tunmer WE (1986). Decoding, reading, and reading disability. Remedial and Special Education, 7, 6–10.

Graesser AC, León JA, & Otero J (2002). Introduction to the psychology of science text comprehension. In Otero J, León JA, & Graesser AC (Eds.), The psychology of science text comprehension (pp. 1–15). Mahwah, NJ: Erlbaum.

Graesser AC, Singer M, & Trabasso T (1994). Constructing inferences during narrative text comprehension. Psychological Review, 101, 371–395. [PubMed: 7938337]

Holmes BC (1987). Children's inferences with print and pictures. Journal of Educational Psychology, 79, 14–18.

Hoover WA, & Gough PB (1990). The simple view of reading. Reading and Writing: An Interdisciplinary Journal, 2, 127–160.

Kaefer T, Neuman SB, & Pinkham AM (2015). Pre-existing background knowledge influences socioeconomic differences in preschoolers' word learning and comprehension. Reading Psychology, 36(3), 203–231. doi:10.1080/02702711.2013.843064

Kendeou P, Bohn-Gettler CM, White MJ, & van den Broek P (2008). Children's inference generation across different media. Journal of Research in Reading, 31, 259–272. doi:10.1111/j.1467-9817.2008.00370.x

Kim Y-S (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. Child Development, 86, 128–144. doi:10.1111/cdev.12293 [PubMed: 25174258]

Kim Y-SG (2016). Direct and mediated effects of language and cognitive skills on comprehension or oral narrative texts (listening comprehension) for children. Journal of Experimental Child Psychology, 141, 101–120. doi:10.1016/j.jecp.2015.08.003 [PubMed: 26342474]

Kim Y-SG (2017). Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER). Scientific Studies of Reading, 21, 310–333. doi:10.1080/10888438.2017.1291643

Kim Y-SG (2020). Hierarchical and dynamic relations of language and cognitive Skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). Journal of Educational Psychology, 112 (4), 667–684. 10.1037/edu0000407

Kim Y-S, & Phillips B (2014). Cognitive correlates of listening comprehension. Reading Research Quarterly, 49, 269–281. doi:10.1002/rrq.74

Kintsch W (1988). The use of knowledge in discourse processing: A construction-integration model. Psychological Review, 95, 163–182. [PubMed: 3375398]

Lapp D, & Flood J (1986). Teaching students to read. New York: Macmillan.

Lepola J, Lynch J, Laakkonen E, Silvén M, & Niemi P (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4- to 6-year old children. Reading Research Quarterly, 47, 259–282. doi:10.1002/RRQ.020

Leslie L, & Caldwell JS (2011). Qualitative reading inventory-5th Edition. Boston, MA: Allyn & Bacon.

Mazany T, Pimentel S, Orr CS, & Crovo M (2015). Reading framework for the 2015 National Assessment of Educational Progress. Retrieved from https://www.nagb.org/content/nagb/assets/documents/publications/frameworks/reading/2015-reading-framework.pdf.

McCormick S (1992). Disabled readers' erroneous responses to inferential comprehension questions: Description and analyses. Reading Research Quarterly, 27, 54–77.

McKenna MC, & Stahl KAD (2009). Assessment for reading instruction – Second Edition. New York, NY: Guilford Press.

McKeown MG, Beck IL, Sinatra GM, & Loxterman JA (1992). The contribution of prior knowledge and coherent text to comprehension. Reading Research Quarterly, 27, 78–93. doi:10.2307/747834

McNamara DS, Kintsch W, Songer NB, & Kintsch E (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. Cognition and Instruction, 14, 1–43. doi:10.1207/sl532690xcil401_l.

McNamara DS, & Magliano J (2009). Toward a comprehensive model for comprehension. In Ross B (Ed.), The psychology of learning and motivation (pp. 297–384). New York, NY: Elsevier.

Meyer BJ, & Ray MN (2011). Structure strategy interventions: Increasing reading comprehension of expository text. International Electronic Journal of Elementary Education, 4(1), 127–152.

Miller SD, & Smith DEP (1985). Differences in literal and inferential comprehension after reading orally and silently. Journal of Educational Psychology, 77, 341–348.

Nation K, Clarke P, Marshall C, & Durand M (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? Journal of Speech, Language, and Hearing Research, 47, 199–211.

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

Oakhill JV, Cain K, & Bryant PE (2003). The dissociation of word reading and text comprehension: Evidence from component skills. Language and Cognitive Processes, 18, 443–468. doi:10.1080/01690960344000008

Oakhill J, Hartt J, & Samols D (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. Reading and Writing: An Interdisciplinary Journal, 18, 657–686. doi:10.1007/s11145-005-3355-z

Pearson PD, & Dole J (1988). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. Champaign, Ill: University of Illinois at Urbana-Champaign.

Pearson PD, & Johnson DD (1978). Teaching reading comprehension. New York, NY: Holt, Rinehart, & Winston.

Perner J, & Wimmer H (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5- to 10-year-old children. Journal of Experimental Child Psychology, 39, 437–471.

Petscher Y, Compton D, Steacy LM, Kinnon H, & McNeish D (2019). Past perspectives and new opportunities for the explanatory item response model. 10.31234/osf.io/vjzet

Potocki A, Ecalle J, & Magnan A (2013). Narrative comprehension skills in 5-year-old children: Correlational analysis and comprehender profiles. The Journal of Educational Research, 106, 14–26. doi:10.1080/00220671.2012.667013

Raphael T (1984). Teaching learners about sources of information for answering comprehension questions. Journal of Reading, 27 (4), 303–311.

Reardon SF, Kalogrides D, Fahle EM, Podolsky A, & Zárate RC (2018). The relationship between test item format and gender achievement gaps on math and ELA tests in fourth and eighth grades. Educational Researcher, 47(5), 284–294. doi: 10.3102/0013189X18762105

Savage R, Cornish K, Manly T, & Hollis C (2006). Cognitive processes in children's reading and attention: The role of working memory, divided attention, and response inhibition. British Journal of Psychology, 97, 365–385. doi:10.1348/000712605X81370 [PubMed: 16848949]

Senechal M, Ouellette G, & Rodney D (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In Neuman SB & Dickinson D (Eds.), Handbook of early literacy research: Vol. 2 (pp. 173–182). New York: Guilford Press.

Stein NL, & Trabasso T (1981). What's in a story: An approach to comprehension and instruction. In Glaser R (Ed.), Advances in instructional psychology (Vol. 2, pp. 213–267). Hillsdale, NJ: Erlbaum.

Strasser K, & del Rio F (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. Reading Research Quarterly, 49, 169–187. doi:10.1002/rrq.68

Sullivan K, Zaitchik D, & Tager-Flusberg H (1994). Preschoolers can attribute second-order beliefs. Developmental Psychology, 30, 395–402.

Tompkins V, Guo Y, & Justice LM (2013). Inference generation, story comprehension, and language in the preschool years. Reading and Writing: An Interdisciplinary Journal, 26, 403–429. doi:10.1007/s11145-012-9374-7

Trabasso T, & Magliano JP (1996). Conscious understanding during comprehension. Discourse Processes, 21, 255–287.

Vacca JL, Vacca RT, Gove MK, Burkey LC, Lenhart LA, & McKeon CA (2009). Reading and learning to read (7th ed.). Boston, MA: Pearson.

Van Dijk TA, & Kintsch W (1983). Strategies of discourse comprehension. New York: Academic Press.

Wechsler D (2009). Wechsler Individual Achievement Test (3rd edition). San Antonio, TX: Pearson.

Wellman HM, Cross D, & Watson J (2001). Meta-analysis of theory-of-mind development: The truth about false belief. Child Development, 72, 655–684. [PubMed: 11405571]

Williams JP, Hall KM, & Lauer KD (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. Exceptionality, 12, 129–144. doi:10.1207/s15327035ex1203_2

Woodcock RW, McGrew KS, & Mather N (2001). Woodcock–Johnson III Tests of Achievement. Itasca, IL: Riverside.

Wolfe MBW (2005). Memory for narrative and expository text: Independent influences of semantic associations and text organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 359–364. doi:10.1348/000709906X143902

Wolfe MBW, & Mienko JA (2007). Learning and memory of factual content from narrative and expository text. British Journal of Educational Psychology, 77, 541–564.doi:10.1348/000709906X143902

Wolfe MBW, & Woodwyk JM (2010). Processing and memory of information presented in narrative or expository texts. British Journal of Educational Psychology, 80(3), 341–362. doi:10.1348/000709910X485700

Zwaan RA, & Radvansky GA (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123, 162–185. [PubMed: 9522683]

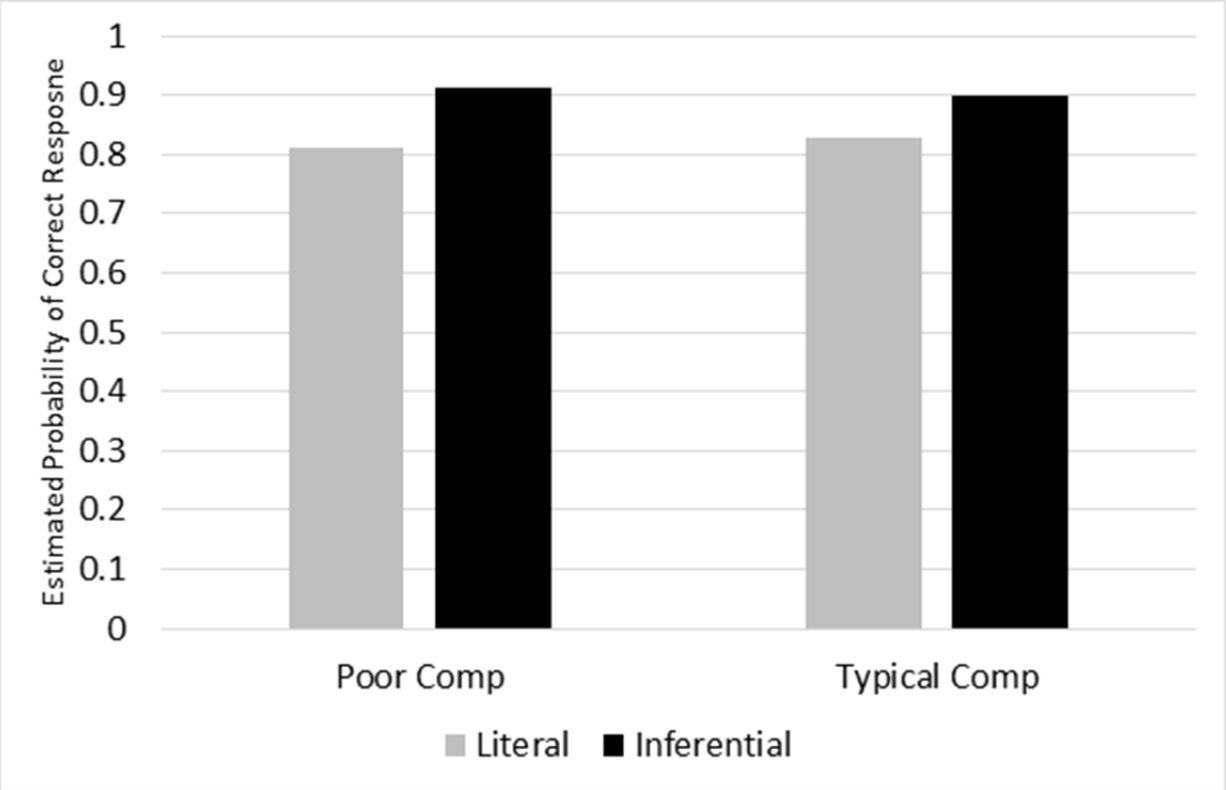

Figure 1.
Interactions between literal and inferential comprehension questions and reading comprehension status (poor and typical) on listening comprehension.

Table 1

Sample means, standard deviations, and correlations among key measures.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Poor Reading Comprehension	0.27	0.44	1.00							
2. Test of Narrative Language: SS	8.47	2.86	−.44	1.00						
3. Qualitative Reading Inventory: Raw	9.35	3.57	−.34	.56	1.00					
4. WJ Picture Vocabulary: SS	96.78	10.29	−.46	.55	.43	1.00				
5. Knowledge-Based Inference: SS	92.70	12.96	−.43	.67	.49	.55	1.00			
6. Theory of Mind: Raw	7.93	4.09	−.27	.54	.51	.40	.48	1.00		
7. Grammaticality Judgement: SS	95.50	13.16	−.52	.62	.46	.58	.66	.40	1.00	
8. Comprehension Monitoring: Raw	6.72	2.97	−.22	.44	.41	.30	.46	.35	.37	1.00
9. Working Memory: Raw	7.78	4.05	−.27	.31	.33	.34	.28	.30	.37	.21

Note. Poor Reading Comprehension = Dichotomous indicator of poor reading comprehension (SS ≤ 85); SS = Standard score; Raw = Raw score; Qualitative Reading Inventory passages (expository texts); WJPV = Woodcock Johnson; All correlations are statistically significant $p < .05$.

Ann Dyslexia. Author manuscript; available in PMC 2022 July 01.

Table 2

Random effect coefficients from unconditional and conditional explanatory item response models

		Random Effects					
Model	**Random Effect**	**Child**	**Classroom**	**School**	**Item**	**Passage**	**Logit**
Unconditional	Variance	0.69	0.14	0.00	2.79	1.47	3.29
	ICC	0.08	0.02	0.00	0.33	0.18	0.39
Model 1	Variance	0.62	0.01	-	2.77	0.03	3.29
	Pseudo-R^2	0.10	0.93	-	0.01	0.98	-
Model 2	Variance	0.62	0.01	-	2.68	0.05	3.29
	Pseudo-R^2	0.10	0.93	-	0.04	0.97	-
Model 3	Variance	0.24	0.01	-	2.68	0.04	3.29
	Pseudo-R^2	0.65	0.93	-	0.04	0.97	-
Model 4	Variance	0.24	0.01	-	2.68	0.05	3.29
	Pseudo-R^2	0.65	0.93	-	0.04	0.97	-

Note. ICC = intraclass correlation. Logit scale variance for all EIRMs is fixed at $\pi^2/3 \sim= 3.29$.

Table 3

Fixed effect coefficients for conditional explanatory item response models

	Model 1			Model 2			Model 3			Model 4		
Fixed Effect	***Est.***	***SE***	***P***	***Est.***	***SE***	***p***	***Est.***	***SE***	***p***	***Est.***	***SE***	***p***
Intercept	1.83	0.54	.001	2.37	0.75	.001	2.17	0.73	.003	2.26	0.74	.003
Expository	−2.37	0.49	<.001	−3.22	0.92	<.001	−3.22	0.90	<.001	−3.33	0.92	<.001
Literal	0.06	0.53	.915	−0.60	0.81	.46	−0.60	0.79	.45	−0.68	0.80	.40
Poor RC	−0.85	0.09	<.001	−0.59	0.15	<.001	0.18	0.15	.229	−0.08	0.07	.29
Exp*Lit				1.20	1.06	.26	1.20	1.04	.25	1.30	1.06	.22
Exp* Poor RC				−0.27	0.16	.10	−0.26	0.16	.10			
Lit* Poor RC				−0.29	0.14	.04	−0.29	0.14	.04			
Exp*Lit*RC				0.27	0.18	.14	0.27	0.18	.14			
Inference							0.02	0.00	<.001	0.02	0.01	.006
Theory of mind							0.07	0.01	<.001	0.05	0.02	.01
Monitoring							0.05	0.01	<.001	0.05	0.03	.06
Vocabulary							0.01	0.00	.001	−0.00	0.01	.96
Grammar							0.01	0.00	.003	0.01	0.00	.04
Working M							0.02	0.01	.004	0.03	0.02	.09
Exp*Infer										−0.01	0.01	.16
Lit*Infer										0.00	0.01	.22
Infer*Exp*Lit										−0.00	0.01	.83
Exp*ToM										0.03	0.02	.14
Lit*ToM										0.01	0.02	.47
ToM*Exp*Lit										−0.03	0.02	.25
Exp*Monitor										0.01	0.03	.83
Lit* Monitor										−0.01	0.03	.60
Monitor *Exp*Lit										0.01	0.03	.74
Exp*Vocab										0.01	0.01	.32
Lit*Vocab										0.01	0.01	.08
Vocab *Exp*Lit										−0.01	0.01	.29
Exp*Gram										−0.00	0.01	.88
Lit*Gram										−0.01	0.01	.40
Gram*Exp*Lit										−0.00	0.01	.89
Exp*WM										0.01	0.02	.77
Lit*WM										−0.02	0.02	.30
WM*Exp*Lit										0.01	0.02	.74

Note. Poor RC = Poor reading comprehension (SS < .85); Exp = Expository passages; Lit = Literal questions; Working M = Working memory; Infer = Inferencing; ToM = Theory of mind; Gram = Grammatical knowledge; WM = Working memory.