Published in final edited form as:

Sci Stud Read. 2014 January 1; 18(6): 415–431. doi:10.1080/10888438.2014.926906.

The Role of Text Memory in Inferencing and in Comprehension Deficits

Anh N. Hua and Janice M. Keenan

University of Denver

Abstract

Comprehension tests often compare accuracy on inferential versus literal questions and find inferential harder than literal, and poor comprehenders performing worse than controls. Difficulties in integration are assumed to be the reason. This research explores another reason – differences in memory for the passage information underlying the questions. Thirty-nine poor comprehenders and 39 controls were given multiple-paragraph passages, which they retold before answering questions. Retellings permitted assessing question accuracy as a function of memory for the text underlying each question. Inferential accuracy was poorer than literal, and the expected group effect obtained. However, when text memory was perfect, group differences disappeared, indicating that poor comprehenders can generate inferences as well as controls, if they have the relevant information in memory. These findings show that text memory is crucial in distinguishing poor comprehension.

Comprehension, especially of longer passages, entails considerable inferential processing, driven by the need to make one's interpretation of the text more coherent and elaborate than the text itself (Keenan, in press). Because inferencing is so central to comprehension, it is a major focus of both assessments of comprehension and accounts of comprehension difficulties. Although inferential processing can be assessed online (using reading times or probe tasks, c.f., Keenan, Potts, Golding, & Jennings, 1990), most commercial tests assess it offline by using inferential comprehension questions presented after the passage. The research in this article is designed to further our understanding of this offline assessment of inference skill by evaluating assumptions that are commonly made when interpreting results based on these questions.

Studies using commercial comprehension tests that contain inferential and literal comprehension questions typically find that performance on inferential questions is poorer than on literal questions (Pearson, Hansen, & Gordon, 1979). The typical explanation for this difference in accuracy assumes differences in computational effort required by the question types. Answering literal questions, which are often phrased as “what” or “how” questions (e.g., *what did the father say?*), is characterized as a process of remembering the passage and matching the queried information to the text information. Answering inferential questions, which are often phrased as “why” questions (e.g., *why did he say that?*), requires

the same processes plus an additional process of integrating across multiple ideas in order to derive an answer. It is this additional integration process that is assumed to make the computations involved in answering inferential questions harder (Perfetti, Landi, & Oakhill, 2005) and to be the reason why poor comprehenders might be particularly poor at drawing inferences (Oakhill, 1982).

Integration is not, however, the only difference in processing between answering literal and inferential questions. There is another difference that is inherent in the computational differences account but often overlooked – it is a difference between literal and inferential questions in the memory load, or the amount of passage information that must be accessed from memory in order to answer the questions. Because inferential questions (such as explaining why something happened) typically require integrating across multiple text elements, they may require the comprehender to remember more of the text in order to come up with the correct answer than what is required to answer literal questions. Perhaps these differences also explain, at least in part, the difficulties associated with answering offline inference questions.

This paper therefore addresses the questions of whether there are differences in how much information needs to be remembered to answer literal and inferential questions in a commonly used commercial test of comprehension, and then assesses the role these differences in text memory might be playing both in the accuracy differences between inferential and literal questions as well as in findings of inferencing deficits in poor comprehension. We examined the questions from the Qualitative Reading Inventory-3, QRI (Leslie & Caldwell, 2001), a test that we have been using as part of a test battery for a larger study of individual differences in comprehension (Keenan, Betjemann, Wadsworth, DeFries, & Olson, 2006). We first checked to see if the QRI inferential questions in fact differ from the literal questions in the amount of text needed to be remembered in order to answer them. Because we found that there is a difference, as we report below, we set out to determine the extent to which it explains performance differences between literal and inferential questions, and the extent to which it underlies the difficulties of poor comprehenders.

Poor comprehenders are often thought to have less ability to integrate text ideas when reading relative to typical readers (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001; Catts, Adlof, & Weismer, 2006; Oakhill, 1982). This can be a result of their reduced working memory (Carretti, Borella, Cornoldi, & De Beni, 2009; Swanson & Berninger, 1995; Yuill, Oakhill, & Parkin, 1989), since WM is needed to maintain the ideas that need to be integrated. It can also be due to poor comprehenders having less knowledge, as reflected in weaker vocabulary (Cain, Oakhill, & Elbro, 2003; Nation, Cocksey, Taylor, & Bishop, 2010; Nation, Clarke, Marschall & Durand, 2004; Nation & Snowling, 1998), which can affect how readily one might integrate ideas together. We are investigating whether, when offline questions are used to assess comprehension, there may be yet another reason why these children may perform poorly: are they less able to access (either for reasons of access or reasons of availability) all the information needed to answer the questions from their long-term memory for the passage (what we are calling *text memory*)? Thus, we examine whether poor comprehenders have more difficulty relative to typical comprehenders on inferential than on literal questions and determine whether

these differences are mirrored in their memory for the question *premises* (the specific parts of the passage needed to answer the questions). Importantly, we also determine whether poor comprehenders continue to show deficits on inference questions when their recall shows that they have all the premises available in memory, or whether having memory for the specific premises is sufficient to correct inference difficulties.

We operationally define text memory in this study as the comprehender's recall of the parts of the passage that are relevant to answering the comprehension questions when they retell the passage. It is conceptually distinct from working memory. Whereas working memory is critical in constructing a mental representation of a passage, text memory refers to the ability to retrieve this mental representation later. Text memory is also distinct from inferencing because comprehenders can retain ideas from a passage without necessarily making connections between them.

In sum, there are two main findings in comprehension research that are typically attributed solely to inference integration processes, but which, because of a confound with memory load, may be due wholly or in part to memory differences. One is the main effect of question type, where accuracy on inferential questions is often poorer than on literal questions. The other is the interaction of comprehension skill with question type, where poor comprehenders sometimes (Cain & Oakhill, 1999; Oakhill, 1982; Oakhill, 1984; Oakhill, Yuill, & Parkin, 1986), but not always (Catts et al., 2006; Spooner, Gathercole, & Baddeley, 2006) show greater deficits relative to controls on inferential than on literal questions.

Previous Research on Text Memory and Inferential Questions

Our study evaluates the contribution of text memory to question type and group difference effects by having participants first retell the passages which they then answer comprehension questions about. This permits us to determine for each question for each participant whether the premises needed to answer a particular question were in their retelling and thus whether they were available in memory. This method of examining question performance contingent on memory allows us to determine to what extent difficulties are due to failures to retrieve versus failures to integrate the information relevant to answering the question.

Our method is an advance over prior research because it is a more direct assessment of availability of the information needed to answer inferential questions. Prior research has tended to use performance on literal questions to assess the role that text memory plays in answering inferential questions. To illustrate, when Cain and Oakhill (1999) interpreted their findings, they reasoned that “less skilled comprehenders were not significantly poorer at answering questions about literal information...Thus, the less skilled comprehenders’ initially poor performance on the inference questions cannot simply be attributed to poor memory for the text per se” (p. 500). The problem is that performance on literal questions only provides a measure of how well the text underlying literal questions is remembered. Because inferential questions often refer to different information in the text, it is quite possible for there to be differences in memorability of the information needed to answer the two types of questions. Thus, it is preferable to have a more direct assessment of the

information underlying inferential questions, like that provided by examining retellings to see if they include the information underlying each question, in order to effectively exclude text memory difficulties as the source of inferential question difficulties.

A review of studies assessing group differences on inferential questions offers some support for the role of text memory in inferential question difficulties. Cain and Oakhill (1999) had skilled comprehenders, less skilled comprehenders, and a comprehension-age match group read four experimental stories and answer six questions about each story immediately after reading it. Less skilled comprehenders did not perform significantly worse on the literal questions than the other groups, but were significantly worse on inferential questions; this led Cain and Oakhill to conclude that a potential mechanism underlying comprehension deficits is impaired inference generation, a conclusion which underlies much current research on poor comprehenders (Cain & Oakhill, 2006; Oakhill & Cain, 2012). Interestingly, however, that same study provided evidence which could be interpreted as suggesting that one of the reasons for poor comprehenders' inference difficulties may be text memory difficulties. When children were allowed to look back at the passages to answer the questions, thereby reducing the memory load, the less skilled comprehenders' performance improved dramatically. It improved not just on the literal questions, but also on the inferential questions, showing that text memory is a factor in answering inferential questions.

Spooner et al. (2006) extended previous studies by Oakhill and colleagues (Oakhill, 1982; Oakhill, Yuill, & Parkin, 1986) which had used endorsement of congruent foils in a recognition test as evidence for inferences. Spooner et al. reduced the number of passages and length of retention interval in order to increase accessibility of text information and found that children's general recognition accuracy improved. However, they did not replicate Oakhill and colleagues' finding of an inference deficit in poor comprehenders; rather, they found that poor comprehenders were just as likely as skilled comprehenders to integrate information (endorse the congruent foils), regardless of the memory demands of the tasks. Furthermore, they proposed that the discrepancy in findings might be due to memory differences in poor comprehenders between the studies because different versions of the Neale Analysis of Reading Ability were used to define comprehension groups. The more recent version used by Spooner et al. allows stories to be available during questioning, whereas the earlier version used by Oakhill and colleagues required the stories to be removed. Thus, poor comprehenders in Oakhill and colleagues' studies may have performed poorly on comprehension questions because text memory difficulties would have contributed to their selection. It raises the possibility that memory factors may play a role in poor performance on inferential questions.

Similarly, Catts et al. (2006) found that even though typical and poor comprehenders both performed more poorly on inference than literal questions, the group difference was not greater for inference questions, again suggesting that poor comprehenders do not have a specific inferencing deficit and that whatever cognitive difficulties underlie inference questions are shared across groups. In addition, Catts et al. (2006) provided further support for the notion that memory difficulties underlie poor performance on inference questions

because inference questions that required integration across adjacent information were more accurate than those integrating distant information.

Finally, support for text memory differences across comprehension skill groups comes from a study by Cain et al. (2001). They were investigating whether group differences in inferencing could be due to knowledge, so they taught the children the knowledge that they would need to draw inferences from a story before they got the story. Although all children mastered the facts from the knowledge base, skilled comprehenders still outperformed poor comprehenders (on literal questions as well as inferential questions), suggesting that group differences in the knowledge needed to integrate the text information is not the only factor explaining inference difficulties. Interestingly, however, there appear to also have been memory differences between the groups in how well the knowledge base was encoded in memory because when tested a week later on retention of the knowledge base, the skilled comprehenders significantly outperformed poor comprehenders. This may have been partly responsible for the difference in group performance on question accuracy. However, it should be noted that when an analysis of covariance was conducted to remove the group difference in memory, poor comprehenders still performed more poorly on inferences. So memory differences are relevant, but they may not be the only factor in poor inferencing for poor comprehenders.

It is evident from this review that questions remain about the role of text memory in inferencing difficulties. Poor comprehenders sometimes show a specific deficit in inferencing (Cain & Oakhill 1999, Cain et al., 2001) and sometimes they do not (Spooner et al., 2006; Catts et al., 2006). Sometimes this can be eliminated by having children look back at the passages (Cain & Oakhill, 1999), but sometimes attempts to equate group memory differences does not eliminate the inference deficit (Cain et al., 2001). The present study does not attempt to manipulate memory factors or equate memory across groups, but instead it examines accuracy on each question directly as a function of whether the question premises were included in the child's retelling. It therefore provides a straightforward assessment of the role that differences in text memory are playing both in the accuracy differences between inferential and literal questions as well as in findings of inferencing deficits in poor comprehenders. As such, it allows us to determine to what extent difficulties are due to failures to retrieve versus failures to integrate the information relevant to answering questions.

Overview of Study

This study uses data from the Qualitative Reading Inventory-3 (Leslie & Caldwell, 2001), a test administered as part of the larger comprehension battery of the Colorado Learning Disabilities Research Center (CLDRC) (Keenan, Betjemann, & Olson, 2008; Keenan, et al., 2006). Because the QRI test provides retellings and comprehension questions on the same passage, it allows us to assess text memory by using the retellings to determine whether the text premises needed to answer the comprehension questions are available in memory.

We compare literal and inferential question accuracy and text memory for both skilled and poor comprehenders. Poor comprehenders are said to have weaknesses in oral language that

impact both listening and reading comprehension (Gough & Tunmer, 1986). We defined our poor comprehender group using listening rather than reading comprehension because we wanted to select readers whose comprehension skills were not due to poor word decoding skills (Gough, Hoover, & Peterson, 1996; Nation & Snowling, 1997). All our poor comprehenders were selected to have a listening comprehension composite score below the 25th percentile. Consistent with previous studies (Cain & Oakhill, 1999; Catts et al., 2006; Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Henderson, Snowling, & Clark, 2013), they were also selected so as to have a *specific* deficit in comprehension. In other words, the poor comprehenders did not have a concomitant deficit in word decoding; their word decoding skill was matched to that of the good comprehenders and was at or above the 50th percentile.

As in other inference studies, we ask: 1) is there a main effect of question type such that inferential questions are more difficult than literal questions, 2) is there simply a main effect of group, or is there a group by question type interaction such that poor comprehenders are particularly poor on inferential questions. In addition, we extend previous studies by also seeing whether the patterns found on question accuracy are mirrored in the premise recall data. Although the number of premises is greater for inference items, if we find that the proportion of premises recalled is less for inferential than literal premises, this would suggest that the question accuracy data are, at least in part, a function of this memory difference. However, if there are no differences in recall of the premises underlying literal and inferential questions, then this would suggest that poor memory is not an explanation for poorer performance on inference questions.

The most novel and important aspect of our results are the contingency analyses that we will also perform. These analyses examine the extent to which accuracy on a particular question depends on memory for its underlying premises. We will examine how often a child recalls all the premises related to the question but then answers the question incorrectly, and determine if that varies as a function of either question type or the child's skill level. If poor comprehenders answer items with the same accuracy as skilled comprehenders when they have perfect memory for the premises, it would suggest that deficits in inferencing may reflect difficulties in memory for the premises and not just integration.

Method

Participants

Thirty-nine poor comprehenders and 39 controls were selected from a sample of 1539 children tested on the language comprehension battery (Keenan et al., 2006; Keenan et al., 2008) of the Colorado Learning Disabilities Research Center (Olson, 2006). Poor comprehenders were selected to have a listening comprehension composite *z*-score defined across the full sample below the 25th percentile, whereas controls had a listening comprehension composite *z*-score at or above the 50th percentile. Poor comprehenders and controls were matched on both age and word reading skill. All children, regardless of group, were selected to have a composite *z*-score on word reading accuracy of 0 or greater; thus, children who experience comprehension difficulties due to poor word reading skills were excluded from the current study. Individual *z*-scores were computed using the standardized

residuals that were saved after regressing each test on children's age and age squared and composite z-scores were obtained by first standardizing each test relevant to the composite and averaging the standardized scores of the relevant tests and then restandardizing the composite scores. Thus, although ages ranged from eight to eighteen years, age did not influence the composite z-scores. Table 1 shows average age, word reading, and listening composite scores of the two groups, and shows that the groups differed significantly only on mean listening comprehension, $t(76) = 18.22, p < .001, d = 4.13$. Children whose full-scale IQ (Wechsler, 1974) was less than 85 were excluded, as were children who had a diagnosis of ADHD.

Group Selection Measures

Listening Comprehension—Two listening comprehension tests were used. One was the Woodcock-Johnson Oral Comprehension subtest (Woodcock et al., 2001) in which participants listen to very short passages and provide a missing word to demonstrate their comprehension. The published test-retest reliability was .88. The other is the KNOW-IT Test (Barnes et al., 1996) where children first learn a novel knowledge base about an imaginary planet and then listen to a long (7-minute) story and answer open-ended comprehension questions. The KNOW-IT Test scores represent accuracy on the comprehension questions, adjusted for having the relevant knowledge base information available in memory.

Word Decoding—Word decoding skill was assessed by the Timed Oral Reading of Single Words (Olson, Forsberg, Wise, & Rack, 1994) and the Peabody Individual Achievement Test (PIAT) word recognition subtest (Dunn & Markwardt, 1970). Reliability estimates were .89 and .95 for the measures, respectively.

Passages and Comprehension Questions

The passages and comprehension questions were part of the Qualitative Reading Inventory-3 (Leslie & Caldwell, 2001). Passages ranged in length from 250 to 785 words and are described in the manual as designed to be grade-level appropriate. Thus, the specific passages given to a child varied with that child's grade in school. The specific passages selected for the current study were: second-grade passages Lucky Cricket and Father's New Game, third-grade passages The Friend and Trip to the Zoo, fourth-grade passages Johnny Appleseed, Amelia Earhart, Cahokia, and Octopus, fifth-grade passages Martin Luther King, Margaret Mead, City of Cahokia, and Octopus, sixth-grade passages Pele, Andrew Carnegie, City of Cahokia, and Octopus, and seventh grade or higher passages Bidy Mason, Malcolm X, Lewis & Clark, and Ferdinand Magellan.

Each passage was accompanied by six open-ended comprehension questions: three literal and three inferential. Literal questions required answers that were stated explicitly in the passage and tended to ask what or how, and inferential questions required an incorporation of a child's prior knowledge such as vocabulary knowledge and schemas about the world situations with text information, and tended to ask why something occurred. Because the passages are so long, Table 2 presents just a portion of a passage and examples of one of the literal and inferential comprehension questions for that passage. As can be seen from this

example, the literal question simply requires the comprehender to remember one fact from the text, whereas the inferential question requires retaining multiple text premises and integrating them on the basis of one's knowledge that "enemies" and "predators" refer to the same thing. The comprehension questions were scored as either correct or incorrect. Trained testers scored children's answers according to the suggested answers provided by the QRI manual (Leslie & Caldwell, 2001). The inter-scorer reliability of the questions was .98.

Premise Identification and Scoring

Because the QRI is a commercial test that labels questions as literal or inferential, but does not specify which premises from the text are relevant to answering each question, we needed to conduct a norming study to identify which information from the text was relevant to each question. Two raters, who had a high familiarity with the test, identified the premises. For each question, they independently established which sentences from the passage went with each question, i.e., needed to be remembered in order to answer the question. The percent of agreement was 87%. Disagreements were resolved through discussion.

Table 2 shows the premises from the text that were identified as underlying the questions. This example is representative of the general trend regarding the number of premises underlying the question types. In this example, there is only one premise underlying the literal question, whereas there are three premises underlying the inferential question. Across all items, the average number of premises was 1.12 for literal and 1.89 for inferential, confirming that there is a greater memory load for inferential than literal. Because of this difference, proportions were therefore used in comparing the amount recalled across literal and inferential questions.

Premise retellings were scored for gist by the first author. In order to ensure that the scoring of premises was reliable, a second rater scored the data from 23 participants (30% of the retellings). The percent of agreement was 90%, and scoring discrepancies between the two raters were resolved through discussion so that scoring was consistent across all participants.

Procedure

Children were given the passages appropriate to their grade level. To conform with the requirements of the larger study for which these data were being collected, half the passages were read aloud and half were listened to, with the assignment of passages to modality the same across all participants. Thus, while we report effects of modality to be complete in our reporting, it is important to keep in mind that modality is confounded with the different passages and thus any modality effects may reflect passage differences as much as modality per se. Similarly, the constraints of the larger study required that all the passages for each grade level be administered in the same order across all participants. Immediately following each passage, the child was asked to retell everything that he/she could remember from the beginning to the end of the passage. Following the retelling, participants answered six open-ended comprehension questions about the passage. Retellings and answers to the comprehension questions were audio-recorded and transcribed, and retellings of question premises were scored as described above.

Results

Correlations of Text Memory and Question Accuracy

There was a high correlation between overall memory for the premises from the text and accuracy in answering the comprehension questions, $r = .82$. This stands in contrast to the rather modest correlation of $r = .41$ that has been reported when QRI retellings are scored, not just for the question premises, but for all the ideas in the passage (Keenan et al., 2008). The difference reflects the fact that there are many more ideas in the passage than those assessed by the six comprehension questions; so when recall is scored for all ideas, different information is being assessed by retellings and question accuracy, whereas when recall is scored specifically for the premises underlying the questions, questions and memory are assessing the same information. The correlation between premise memory and question accuracy was high for both the controls ($r = .85$) and the poor comprehenders ($r = .72$), and the difference in correlations was not significant, $z = 1.48$, $p = .14$. The high correlations between text memory and question accuracy demonstrates the interconnectedness of text memory and question accuracy in both groups.

Comprehension Question Accuracy

A three-factor analysis of variance¹ was performed with comprehension group (poor comprehenders vs. controls) as a between-subjects factor, and question type (literal vs. inferential) and modality (listening vs. reading) as within-subjects factors. Consistent with prior research, a main effect of question type emerged: accuracy was greater for literal than for inferential questions, $F(1, 76) = 69.18$, $p < .001$, partial $\eta^2 = .48$. A main effect of group was also found: Controls performed significantly better than poor comprehenders, $F(1, 76) = 23.81$, $p < .001$, partial $\eta^2 = .24$. As can be seen in Figure 1, there was not a significant interaction between comprehension group and question type, $F(1, 76) < 1$. Thus, while poor comprehenders performed worse on inferential questions than on literal questions, the same pattern was found for the controls. For simplicity, Figure 1 collapses across modality. The main effect of modality was not significant, $F(1, 76) < 1$. Nor did modality significantly interact with any factors; modality X group $F(1, 76) < 1$; modality X question type, $F(1, 76) = 2.27$, $p = .14$; and question type X group X modality $F(1, 76) = 1.97$, $p = .17$.

Text Memory

A similar three-factor analysis of variance was performed on memory for the question premises. Figure 2 presents the results; again, we collapsed over modality because all interactions were nonsignificant, with $F_s < 1$. In contrast to question accuracy, where performance on literal was significantly greater than on inferential, we found no difference in proportion recalled of the premises related to inferential and literal questions. Memory for the premises underlying literal questions was only slightly better (63%) than for the premises underlying inferential questions (60%), $F(1, 76) = 1.88$, $p = .17$. As with question accuracy, we again see a main effect of group, with controls ($M = .69$, $SD = .14$) recalling

¹Although the question accuracy data did not meet the assumptions of normality and homogeneity of variance, the violations were not severe enough to affect the results because ANOVA is quite robust to the violation of normality for a sample size as large as ours (Keppel & Wickens, 2004), and because the main effects of group remained significant even when equal variances were not assumed, $t(63.15) = 4.88$, $p < .001$ for question accuracy.

more premises than poor comprehenders ($M = .54$, $SD = .18$), $F(1, 76) = 17.37$, $p < .001$, partial $\eta^2 = .19$. The interaction between premise question type and group, as can be seen in Figure 2, was not significant, $F(1, 76) < 1$. The main effect of modality was significant with performance on reading passages slightly better than that on listening passages, $F(1, 76) = 4.18$, $p = .04$, partial $\eta^2 = .05$; but as noted in the method, this could reflect passages differences as much as modality differences because modality is confounded with the specific passages.

Although the proportion of premises recalled shows the same group effect as the question accuracy data suggesting that poor comprehenders have poorer text memory relative to controls, the figures show that the patterns across question type are different for the two variables. Whereas the question accuracy results show that literal items are significantly better than inferential, we find that memory for the premises is about as good for premises relevant to inferential questions as it is for those underlying literal questions. This suggests that whatever is responsible for poorer performance on inferential than literal questions may not be simply poor text memory. To further clarify this, we examined the contingency data so that we could determine how often children answered the questions correctly when their retellings showed that they had all the premises available in memory.

Question Accuracy Contingent on Complete Text Memory

Table 3 presents the contingency data. To provide a context for these data, the first row shows for each skill group the average question accuracy, regardless of whether the relevant premises were recalled. The second row shows the contingent question accuracies; in other words, it shows how often each skill group answered the question correctly when they had recalled all the premises related to the question.

By definition, answering literal questions should only require that the child remember what was said in the text in order to get the question correct. Thus, we might expect to find that for literal questions, question accuracy when it is contingent on the child having included the premise in his/her recall, should be quite high. Indeed, 95% of the time when controls had perfect recall for the premises, they answered literal questions correctly. Most interestingly, poor comprehenders answered at a similarly high rate, 93%. This high rate of performance for poor comprehenders on literal questions is in contrast to their 73% when accuracy is noncontingent. It shows that when they have the information in memory, they can answer literal questions just as well as skilled comprehenders. Thus, it suggests that one of the major differences between skilled and poor comprehenders is in their memory for the text information underlying questions.

Answering inferential questions requires more than just remembering; these questions require integrating the text information together with one's prior knowledge to formulate an answer. Thus, in contrast to the literal questions, even when all the premises underlying an inferential question have been remembered, question accuracy may not be as high. In fact, when children had perfect recall for all the premises underlying inferential questions, 82% of the time controls answered the question correctly compared to 77% of the time for poor comprehenders.

In order to determine if this difference in question accuracy between poor and skilled comprehenders is significant, we performed a two-factor ANOVA² with comprehension group (controls and poor comprehenders) as a between-subjects factor and question type (literal and inferential) as a within-subjects factor. If the difference in inferential question accuracy is significant, even when poor comprehenders have all of the premises available in memory, it would suggest that they have a deficit in integration. However, the results showed that neither the main effect of skill group on question accuracy, $F(1,72) = 1.96, p = .17$, nor the interaction between question type and skill group was significant, $F(1, 72) < 1$. Furthermore, a planned comparison *t*-test examining the group difference for inferential questions was not significant, $t(72) = 1.11, p = .27$. The only significant effect was question type; even when text memory was perfect for both types of questions, performance on literal questions was still superior relative to performance on inferential questions, $F(1,72) = 29.39, p < .001$, partial $\eta^2 = .29$.

Discussion

Inferencing is a major focus of both assessments of comprehension and accounts of comprehension difficulties. Most commercial tests assess it offline by using inferential comprehension questions presented after the passage, and find that performance on inferential questions is poorer than on literal questions (Pearson, Hansen, & Gordon, 1979). Poorer performance on inference questions can be attributed to the need to perform an integration process across ideas in order to answer inferential questions - a process that is thought to be a specific deficit in poor comprehenders (Oakhill, 1982; Cain & Oakhill, 2007).

The purpose of the research presented here was to explore the extent to which the above findings on inferential questions might also be attributable to another difference between literal and inferential questions – a difference in the amount of text information that must be retained in memory in order to answer inferential questions. Inference questions often require integrating multiple units of text, and indeed when we examined the questions on the QRI (Leslie & Caldwell, 2001), we found that inferential questions did require remembering more of the text to answer correctly than literal questions. We then asked whether this increased text memory load might help explain why inferential questions are often harder than literal questions, and why poor comprehenders might do more poorly on inferential questions.

We took advantage of the fact that the QRI has children retell each passage before answering the comprehension questions, and used the retellings to determine whether the passage information needed to answer each question was available in the child's memory. These contingency analyses allowed us to determine whether poor comprehenders would still have difficulties with inference questions even when their retelling showed that they had

²Because the analysis requires that each child must have recalled at least both a full premise underlying a literal and an inferential question, any child that did not meet this criterion (i.e. by recalling either a partial or no premise for all the questions or recalling a full premise for one type of question and not the other) was excluded from the analysis. Because two poor comprehenders did not meet this criterion, they were excluded along with their matched controls. Overall, controls recalled a full premise for 63% of the total number of comprehension questions. Poor comprehenders, on the other hand, recalled a full premise for 49% of the total number of questions.

all the premises available in memory, or whether having memory for the specific premises was sufficient to correct inference difficulties.

Our contingency analyses yielded a surprising finding in view of the literature suggesting that poor comprehenders have specific deficits in inferencing (e.g., Cain & Oakhill, 1999; Cain & Oakhill, 2007; Cain et al. 2001). They showed that poor comprehenders did not have any greater difficulty with inference questions than controls when they had the question premises available in memory. This was in contrast to when we examined question accuracy in the usual way (i.e., not contingent on having recalled the question premises), and showed the typical finding of poor comprehenders performing below controls. The fact that the group effect can be eliminated when poor comprehenders have memory for the premises underlying the questions shows that poor text memory can be quite an important factor in distinguishing poor comprehension. For these questions, poor comprehenders had no difficulties in integration. This finding raises the question of whether previous reports of inference deficits in poor comprehenders might be due more to text memory deficits than integration deficits. This possibility needs to be examined in future research. However, we think there are good reasons to presume that text memory is only part of the reason why poor comprehenders can show deficits on inference questions, not the full story.

Our thinking on this question recognizes that there is a selection process controlling which premises are remembered. In other words, which premises are remembered is not randomly assigned but is a function of the comprehender's knowledge and skills. For example, if a portion of the text is familiar, funny, or interesting, then that portion is likely to be better remembered. It seems reasonable to think that the factors that govern which premises are completely recalled are the very same factors that would make integration easier. Thus, the very information that is easiest to remember is the easiest to integrate. Our finding that poor comprehenders do not show deficits on inference questions when they have all the premises in memory therefore means that they can integrate when their interests and knowledge choose the premises to be integrated. But that does not mean that they would not show integration deficits under other conditions.

Another of our findings also suggested that poor comprehenders do not have a specific deficit in inferencing. Specifically, we found no significant interaction between question type and skill group in either the overall or the contingent analyses. This is similar to the findings of Catts et al. (2006) and Spooner et al. (2006), and consistent with claims made by others that poor comprehenders may not have a specific difficulty with inference generation (Adams, Bell, & Perfetti, 1995; Keenan, 2002). It does, however, conflict with studies that do find an interaction (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001; Oakhill, 1982).

It is difficult to know why some studies obtain a question type by skill group interaction while other studies do not. There are many factors that can vary across studies that could influence the effect. One factor that can produce an interaction is if the accuracy level of the control group is at or near ceiling; in that case, differences between literal and inferential questions may only be observed for the poor comprehenders because they are not at ceiling, and thus an interaction is observed. Also, different studies use different passages and

questions, and it may be that the vocabulary knowledge or other types of prior knowledge needed to integrate text information to answer the inferential questions might differ between groups. There are also differences across studies in how the groups are selected; as noted in the introduction, that is how Spooner et al. (2006) suggested reconciling differences between their findings and Oakhill's (1982) regarding the interaction of question type and skill group.

Keenan and colleagues (Keenan et al., in press; Keenan & Meenan, 2014) have recently shown that differences in the selection criteria, as well as the specific tests used to define poor comprehension, lead to varying cognitive profiles in poor comprehenders. It is quite possible that these cognitive differences may be responsible for some of the differences across studies. However, it is not evident what if any particular factor may be responsible for finding that poor comprehenders are specifically poor at inferential questions. One may wonder if using a composite of listening comprehension tests to determine who was a poor comprehender, as we did in this study, was responsible for finding no interaction between question type and group. However, other studies that used reading, instead of listening, comprehension have also found no evidence for a specific inference deficit in poor comprehenders (Catts, et al., 2006; Spooner, et al., 2006).

In addition to examining the role that text memory may play in explaining differences between poor and skilled comprehenders, our results also assessed its role in explaining the question type effect - why inferential questions are often harder than literal questions. Although inferential questions require remembering more of the text, we found that the proportion of premises recalled showed no difference between literal and inferential, for either group. In addition, in the contingency analyses, we found that even when text memory was complete, inferential questions were still answered significantly less accurately than literal questions. It thus appears that what makes inferential questions harder is not just the memory load but the computational complexity of the integration processes required to answer inference questions (Perfetti et al., 2005).

Our experiences in testing have led us to consider another possibility that may underlie some integration failures. Consider this inferential question from one of our passages: "Why might the shy octopus attack another creature?" Many children had recalled the premises that underlie this question: "Crabs are its favorite food" and "The octopus bites into the crab with its strong beak." However, when presented with the question, they would answer it incorrectly by either saying "because it's scared" or "for self-defense." It is as if the words *shy* and *attack* in the question steer them to their own experiences, leading them away from the text premises. Potts and Peterson (1985) referred to such failures to use relevant knowledge that is clearly available in memory as knowledge compartmentalization. Compartmentalization demonstrates that while having adequate text memory is necessary, it may not be sufficient to answering inferential questions.

The present study used retellings of the passages to measure whether the question premises were available in the children's memory. A limitation of using retelling as a baseline for text memory is that it is a complex skill that relies on other skills such as working memory (Kendeou, Papadopoulos, & Spanoudis, 2012). Thus, it will be important for future

research to examine whether similar results obtain when text memory is assessed using measures that may be less reliant on working memory, such as recognition of text premises.

In sum, the current research captures the complex relationship between comprehension skill, text memory, and inferencing. By assessing text memory directly and independently from inferential question accuracy, we found poor comprehenders to be just as capable as controls at integrating text premises when they had retained the text premises. This suggests that their poor performance on inference questions does not always reflect a failure to integrate, but can stem from text memory difficulties. Distinguishing these sources of poor comprehension is an important goal because of the implications for how we should structure assessment, intervention, and remediation (Keenan, 2014). Children whose poor comprehension stems from integration difficulties may benefit more from remediation focused specifically on integration, such as learning to recognize textual cues that signal an inference, whereas children whose poor comprehension stems primarily from text memory difficulties may benefit more from memory training. The current findings suggest that the way to move forward is to disentangle text memory from integration in assessing inferential comprehension using question accuracy.

Acknowledgments

This research was supported by grant P50 HD27802 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to the Colorado Learning Disabilities Research Center, for which J. Keenan is a co-PI.

References

- Adams BC, Bell LC, Perfetti CA. A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*. 1995; 20:307–323.
- Barnes MA, Dennis M, Kaefele-Kalvaitis J. The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from fifteen years of age. *Journal of Experimental Child Psychology*. 1996; 61:216–241. [PubMed: 8636665]
- Bowyer-Crane C, Snowling MJ. Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*. 2005; 75:189–201. [PubMed: 16033662]
- Cain K, Oakhill J. Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*. 1999; 11:489–503.
- Cain K, Oakhill J. Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*. 2006; 76:683–696. [PubMed: 17094880]
- Cain, K.; Oakhill, J. Reading comprehension difficulties: correlates, causes and consequences.. In: Cain, K.; Oakhill, J., editors. *Children's comprehension problems in oral and written language: a cognitive perspective*. the Guilford Press; NY: 2007. p. 41-76.
- Cain K, Oakhill J, Barnes MA, Bryant PE. Comprehension skills, inference making ability, and their relation to knowledge. *Memory and Cognition*. 2001; 29:850–859. [PubMed: 11716058]
- Cain K, Oakhill JV, Elbro C. The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of child language*. 2003; 30:681–694. [PubMed: 14513473]
- Carretti B, Borella E, Cornoldi C, De Beni R. Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*. 2009; 19:246–251.
- Catts HW, Adlof SM, Weismer SE. Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*. 2006; 49:278–293.

- Dunn, LM.; Markwardt, FC. Examiner's manual: Peabody individual achievement test. American Guidance Service; Circle Pines, Minnesota: 1970.
- Elwér Å, Keenan JM, Olson RK, Byrne B, Samuelsson S. Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*. 2013; 115:497–516. [PubMed: 23528975]
- Gough, PB.; Hoover, WA.; Peterson, CL. Some observations on the simple view of reading.. In: Cornoldi, C.; Oakhill, J., editors. *Reading comprehension difficulties: Processes and intervention*. Erlbaum; Mahwah, NJ: 1996. p. 1-14.
- Gough PB, Tunmer WE. Decoding, reading, and reading disability. *Remedial and special education*. 1986; 7:6–10.
- Henderson L, Snowling M, Clarke P. Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*. 2013; 17:177–198.
- Keenan, JM. Inferences.. In: Smelser, N.; Baltes, P., editors. *International Encyclopedia of the Social & Behavioral Sciences*. Vol. 11. Pergamon; Amsterdam: 2002. p. 7432-7435.
- Keenan, JM. Psychology of inferences. In: Wright, JD., editor. *International Encyclopedia of the Social & Behavioral Sciences*. 2nd Edition. Elsevier; Oxford, UK: in press
- Keenan, JM. Assessment of reading comprehension.. In: Stone, CA.; Silliman, ER.; Ehren, BJ.; Wallach, GP., editors. *Handbook of Language and Literacy: Development and Disorders*. 2nd Edition. Guilford Press; New York: 2014.
- Keenan JM, Betjemann RS, Olson RK. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12:281–300.
- Keenan JM, Betjemann RS, Wadsworth SJ, DeFries JC, Olson RK. Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading*. 2006; 29:79–91.
- Keenan, JM.; Hua, AH.; Meenan, CE.; Olson, RK.; Pennington, BF.; Willcutt, EG. Issues in identifying poor comprehenders.. *L'année Psychologique/Topics in Cognitive Psychology*. in press
- Keenan JM, Meenan CE. Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*. 2014; 47:125–135. [PubMed: 22442251]
- Keenan, JM.; Potts, GR.; Golding, JM.; Jennings, TM. Which elaborative inferences are drawn during reading?: A question of methodologies.. In: Balota, D.; Flores d'Arcais, G.; Rayner, K., editors. *Comprehension processes in reading*. Erlbaum; Hillsdale, NJ: 1990. p. 377-402.
- Kendeou P, Papadopoulos TC, Spanoudis G. Processing demands of reading comprehension tests in young readers. *Learning and Instruction*. 2012; 22:354–367.
- Kintsch, W.; Kintsch, E. Comprehension.. In: Paris, SG.; Stahl, SA., editors. *Children's Reading Comprehension and Assessment*. Erlbaum; Mahwah, New Jersey: 2005. p. 71-92.
- Leslie, L.; Caldwell, J. *Qualitative Reading Inventory-3*. Addison Wesley Longman; New York: 2001.
- Nation K, Clarke P, Marshall CM, Durand M. Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research*. 2004; 47:199–211.
- Nation K, Cocksey J, Taylor J, Bishop D. A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*. 2010; 51:1031–1039. [PubMed: 20456536]
- Nation K, Snowling. Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*. 1997; 67:359–370. [PubMed: 9376312]
- Nation K, Snowling M. Semantic Processing and the Development of Word-Recognition-Skills: Evidence from Children with Reading Comprehension Difficulties. *Journal of Memory and Language*. 1998; 39:85–101.
- Oakhill J. Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology*. 1982; 73:13–20. [PubMed: 7059748]
- Oakhill J. Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*. 1984; 54:31–39.

- Oakhill JV, Cain K. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*. 2012; 16:91–121.
- Oakhill J, Yuill N, Parkin A. On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading*. 1986; 9:80–91.
- Olson RK. Genes, environment, and dyslexia: The 2005 Norman Geschwind memorial lecture. *Annals of Dyslexia*. 2006; 56:205–238. [PubMed: 17849199]
- Olson, R.; Forsberg, H.; Wise, B.; Rack, J. Measurement of word recognition, orthographic, and phonological skills.. In: Lyon, GR., editor. *Frames of reference for the assessment of learning disabilities: New views on measurement issues*. Brookes; Baltimore: 1994. p. 243-277.
- Pearson PD, Hansen J, Gordon C. The effect of background knowledge on young children's comprehension of literal and inferential information. *Journal of Reading Behavior*. 1979; 11:201–209.
- Perfetti, CA.; Landi, N.; Oakhill, J. The acquisition of reading comprehension skill.. In: Snowling, MJ.; Hulme, C., editors. *The science of reading: A handbook*. Blackwell; Malden, MA: 2005. p. 227-247.
- Potts GR, Peterson SB. Incorporation versus compartmentalization in memory for discourse. *Journal of Memory and Language*. 1985; 24:107–118.
- Spooner ALR, Gathercole SE, Baddeley AD. Does weak reading comprehension reflect an integration deficit? *Journal of Research in Reading*. 2006; 29:173–193.
- Swanson HL, Berninger V. The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*. 1995; 21:83–108.
- Wechsler, D. *Examiners' manual: Wechsler Intelligence Scale for Children-Fourth Edition*. Psychological Corporation; San Antonio, TX: 1974.
- Woodcock, RW.; McGrew, KS.; Mather, N. *Woodcock-Johnson III tests of achievement*. Riverside; Itasca, IL: 2001.
- Yuill NM, Oakhill JV, Parkin AJ. Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*. 1989; 80:351–361. [PubMed: 2790393]

Figure 1.
Question accuracy on literal and inferential questions.

Figure 2.
Proportion of text premises recalled that were relevant to literal and inferential questions.

Table 1

Mean Z-Scores (Standard Deviations) on Group Selection Variables: Age (Years), Word Reading, and Listening Comprehension

	Age	Word reading	Listening comprehension
Controls	11.64 (2.69)	.71 (.39)	.69 (.46)
Poor comprehenders	11.64 (2.67)	.69 (.39)	−1.23 (.47)

Table 2

Excerpt of a Passage from the Qualitative Reading Inventory-3 (Leslie & Caldwell, 2001) and Example of Premises for a Literal and an Inferential Question

Some people think of the octopus as a giant creature. They have seen this in science fiction movies. They think the octopus is a mean creature who attacks people and other animals. The octopus is really a shy animal.

The octopus has eight arms. Its name tells us this because “octo” means eight. The octopus uses its arms to walk on the ocean floor. Its arms are also used to capture crabs. Crabs are its favorite food. The octopus bites into the crab with its strong beak. This sends poison into the crab's body.

Suppose the octopus sees an enemy. The octopus will completely change color. Then it seems to disappear into the background of its hiding place. It is hard for the predator to find the octopus.

• **Literal Question:** What is the favorite food of the octopus? (**Answer:** crabs)

Relevant Text Premise: (1) Crabs are its favorite food.

• **Inferential Question:** Why is it hard for predators to find an octopus? (**Answer:** because it camouflages itself with the background)

Relevant Text Premises: (1) Suppose the octopus sees an enemy. (2) The octopus will completely change color. (3) Then it seems to disappear into the background of its hiding place.

Table 3

Percent Accuracy Overall and When Full Premise (Standard Deviations) Was Recalled

	Literal		Inferential	
	Controls	Poor comprehenders	Controls	Poor comprehenders
Overall accuracy	90 (10)	73 (19)	75 (16)	59 (20)
Contingent accuracy	95 (8)	93 (11)	82 (17)	77 (26)