Comprehension of Inferences in Discourse Processing by Adolescents With and Without Language Impairment

Courtney Karasinski and **Susan Ellis Weismer**
University of Wisconsin—Madison

Abstract

Purpose—This study investigated inference construction within spoken narratives in adolescents with varying cognitive and language abilities, using W. Kintsch's (1988) construction-integration model as a framework. The role of working memory in inference construction was examined along with language and nonverbal cognition.

Method—Participants were 527 eighth-grade students in 4 diagnostic groups: normal language (NL), low cognitive (LC), specific language impairment (SLI), and nonspecific language impariment (NLI). Participants answered premise and inference questions based on adjacent and distant information.

Results—Distant inferences were significantly more difficult than were adjacent inferences. When controlling for premise accuracy, the NL group performed significantly better than each of the other groups on distant inferences. The LC group demonstrated significantly higher accuracy on distant inferences than did the NLI group. Regression analyses revealed that performance on a verbal working memory measure predicted unique variance in distant inference accuracy beyond that accounted for by measures of language and nonverbal cognition.

Conclusions—Understanding implicit information, particularly when linking distant information, is difficult for adolescents who are deficient in language comprehension, verbal working memory skills, and/or general world knowledge.

Keywords

inference; working memory; specific language impairment; comprehension

Production and comprehension of narrative discourse predicts success in daily communication better than isolated language tasks, which do not encompass communicative intent (Humphries, Cardy, Worling, & Peets, 2004). Constructing inferences can facilitate the coherent representation of discourse that is necessary for comprehension (Cain, Oakhill, & Bryant, 2001; Virtue, Haberman, Clancy, Parrish, & Beeman, 2006; Virtue & van den Broek, 2004; Virtue, van den Broek, & Linderholm, 2006). Furthermore, understanding inferences is important in order for adolescents to be successful both in school and in social situations, as inferences are used commonly in daily interactions (Moran & Gillon, 2005).

Kintsch's (1988) construction-integration model of discourse processing provides an explanation of how inferences are formed. Kintsch posited that inferencing occurs in two phases, the first phase consisting of constructing inferences and the second phase consisting of integrating the inferences into a coherent text base. The construction phase was described

as having four steps. In the first step, linguistic input and general knowledge combine to form a proposition or concept. Next, each concept or proposition activates related information. The information with the most activation goes into working memory and an inference is generated. The third step involves creating additional inferences, or counterexamples. At times, the creation of the additional inferences requires active concentration on the information in working memory in order to find additional information. The final step in the construction of inferences involves the interconnections between the inferences and general knowledge. In the integration phase, the text base is integrated into a "coherent whole" when activation stabilizes. This model suggests that comprehension of linguistic input, general world knowledge, and working memory are essential for the construction of inferences.

Given the complex interplay among comprehension of linguistic input, general world knowledge, and working memory, it is not surprising that difficulty in making inferences when processing spoken and written discourse has been documented in individuals with a variety of disabilities, including traumatic brain injury (Moran & Gillon, 2005), language-learning disabilities (Crais & Chapman, 1987), specific language impairment (SLI; Bishop & Adams, 1992; Ellis Weismer, 1985), nonverbal learning disabilities (Humphries et al., 2004; Worling, Humphries, & Tannock, 1999), and reading disabilities (Snyder & Downey, 1991), specifically, poor reading comprehension (Catts, Adolf, & Ellis Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004; Yuill, Oakhill, & Parkin, 1989).

Distance and Working Memory

The distance between the information needed to generate an inference has been shown to play a role in whether or not readers or listeners accurately construct the inference in studies of typical adults (Lea, Mulligan, & Walton, 2005), adolescents with traumatic brain injury (Moran & Gillon, 2005), and adolescents and school-age children with poor reading comprehension (Catts et al., 2006; Yuill et al., 1989). The ability to link widely spaced information to construct an inference correlates with working memory (Catts et al., 2006, Moran & Gillon, 2005). Yuill et al. (1989) attributed the increased ability of good (reading) comprehenders to make inferences to their ability to hold more information in working memory while processing new information. Further support for the role of working memory in inference construction comes from Virtue, Haberman, et al.'s (2006) investigation, which demonstrated that individuals with high working memory were more likely than those with low working memory to make the inference when it was necessary for establishing coherence. On the basis of their neuroimaging (i.e., f MRI) findings, these investigators also suggested that differences in the location of activation may indicate that those with high working memory had made the inference and were integrating the information, whereas those with low working memory were still constructing the inference.

Inference Abilities in Children With Language Impairment

Children with SLI have intact general world knowledge but exhibit difficulty with language comprehension (Bishop & Adams, 1992). These children also have been shown to have deficits in performance on working memory tasks (Ellis Weismer, Evans, & Hesketh, 1999; Marton & Schwartz, 2003; Montgomery, 2002; Montgomery & Evans, 2009) and to display atypical neural substrates for verbal working memory (Ellis Weismer, Plante, Jones, & Tomblin, 2005). Mainela-Arnold and Evans (2005) found that group differences between children with SLI and age-matched typically developing peers in the percentage of words recalled on the competing language processing task (CLPT; Gaulin & Campbell, 1994) were no longer significant when receptive or expressive language scores were controlled, supporting the view that verbal working memory capacity and language knowledge are not

separable entities (MacDonald & Christiansen, 2002). However, this conclusion needs to be viewed cautiously in light of the fact that the small sample size ($n = 10$ per group) may have contributed to the lack of significant group differences once the overlapping variance was removed.

The decreased performance of children with SLI in the areas of language comprehension and/or verbal working memory may contribute to their deficits on tasks involving the construction of inferences. Ellis Weismer (1985) found that children with SLI performed worse than same-age controls who were matched for nonverbal cognition on both premise and inference questions. A conditional probability analysis revealed that even when the children with SLI did accurately recall the premise information, they still did not generate as many correct inferences as the children in the control group; instead, the performance of the children with SLI was similar to that of younger children who were matched for language skills. The findings of Crais and Chapman (1987) indicated that answering inference questions was more difficult than answering premise questions for children with language-learning disabilities, children in an age-matched control group, and younger children who were matched for receptive vocabulary skills. The children with language-learning disabilities did not perform as well on either premise or inference questions as the age-matched control group and did not perform significantly differently from the receptive vocabulary–matched group of younger children. Bishop and Adams (1992) reported that children with SLI performed worse than age-matched controls and similarly to younger language-matched children on answering premise and inference questions about stories presented verbally or pictorially. They suggested that literal questions can be as difficult as questions requiring the generation of inferences when children fail to impose structure on a story by using constructive processing to connect the ideas in the text, making it more difficult to recall the details.

Humphries et al. (2004) and Worling et al. (1999) compared the ability of children with nonverbal learning disabilities (NLDs) and verbal impairment (VI) with a control group of typically developing peers to answer inference questions about orally presented stories. The children were assigned to their diagnostic groups on the basis of discrepancies between performance IQ and verbal IQ and discrepancies between arithmetic scores and reading or spelling scores on an achievement test. It is important to note that the children in the VI group in these studies did not necessarily have SLI. The higher of the two IQ scores was above 85, but the lower score was not necessarily in the range of impairment. Additionally, verbal impairments could have been in areas of reading or spelling rather than spoken language. Humphries et al. (2004) found significant differences between children in the NLD group and the control group in answering inference questions, but neither group differed significantly from the VI group in answering inference questions. Worling et al. (1999) also reported no differences between the NLD and VI groups. The control group performed better than the NLD group on spatial and emotional inferencing tasks, better than the VI group on a general inferencing task, and better than both the NLD and VI groups on a standardized inferencing subtest.

Purpose of the Present Study

The purpose of the present study was to investigate inference construction within spoken narratives in a large population-based sample of adolescents who had widely varying cognitive and language abilities, using Kintsch's construction-integration model as a framework. The ability of adolescents with SLI to construct inferences on the basis of closely or widely spaced information was compared with that of adolescents with normal language (NL), low nonverbal cognition (LC), and low nonverbal cognition with language impairment (nonspecific language impairment; NLI). These groups varied along key

features posited to be essential for inference construction according to Kintsch's model— namely, comprehension of linguistic input, general world knowledge, and working memory. This investigation sought to examine the impact of differing profiles of skills on inference performance. Although previous studies have demonstrated inference difficulties in school-age children with SLI, it is not clear whether these problems persist into adolescence. Prior research has not manipulated the distance between the premises on which the inferences were based or evaluated the role of working memory in the construction of inferences by individuals with SLI. Clinically based convenience samples, rather than population-based samples, are typically used in investigations of individuals with SLI. Clinically based samples may not include a representative number of participants with mild to moderate impairments and may overrepresent individuals who present with co-occurring conditions. Such samples could result in an overestimation of the difficulties exhibited by individuals with SLI (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004). The present study has the advantage of using a recent population sample of school-age children and adolescents with language impairment residing in the United States.

The following research questions were examined in this study: (a) Are inference questions based on distant information more difficult to answer than inference questions based on adjacent information? (b) Do adolescents without language impairment answer distant inference questions with greater accuracy than adolescents with language impairment? (c) Are there differences in the types of errors across groups? and (d) Does working memory performance predict variation in distant inference accuracy beyond that explained by language and nonverbal IQ?

Method

Participants

Participants were 527 eighth-grade students with a mean age of 13.92 years ($SD = 0.40$). The participants in this study were drawn from a stratified cluster sample of children who initially participated in an epidemiologic study in which language impairments in kindergartners were investigated (Tomblin et al., 1997). After the epidemiologic study had been completed, a subsample of children was retained to participate in a longitudinal study. This use of a population-based sample decreases the likelihood of magnifying the deficits of individuals with language impairment by including children with mild to moderate deficits and avoiding the overrepresentation of adolescents who also have co-occurring conditions in addition to language impairment (Fey et al., 2004). Students who did not speak English as their primary language or who had a history of mental retardation, autism, neurological impairment, visual impairment, or hearing loss were excluded from participation in the diagnostic testing phase of the study.

When the students were tested in eighth grade, they were assigned to one of four diagnostic groups on the basis of the EpiSLI system described in Tomblin, Records, and Zhang (1996). The language assessment battery was composed of the Peabody Picture Vocabulary Test— Revised (PPVT–R; Dunn & Dunn, 1981), the Expressive scale of the Comprehensive Receptive and Expressive Vocabulary Test (CREVT; Wallace & Hammill, 1994), Concepts and Directions and Recalling Sentences subtests of the Clinical Evaluation of Language Fundamentals, Third Edition (CELF–3; Semel, Wiig, & Secord, 1995), and a measure of narrative comprehension and production using passages from the Qualitative Reading Inventory–(Leslie & Caldwell, 2001). The scores on each of the language tests were converted to z scores using norms from the larger Iowa sample that used differential weightings adjusting for the larger sample of children with LI relative to their prevalence in the general population. Scores on the complete language battery were converted to a composite z score in which each of the measures was weighted equally. Students who

obtained two or more of the five language *z* scores that fell more than 1.25 standard deviations below age norms were considered to have LI. Performance intelligence quotient (PIQ) was measured using the Block Design and Picture Completion subtests of the Wechsler Intelligence Scale for Children—Third Edition (WISC–III; Wechsler, 1991). Students who earned a PIQ below 87 were considered to have low cognitive abilities. The NL control group consisted of the students whose composite scores on the language measures and nonverbal cognition measures fell within the normal range. The students whose scores fell into the range of impairment on the language measures but within the normal range on the WISC–III comprised the group with SLI. The students scoring below 87 on the WISC–III but in the normal range on the language measures were in the LC group. The NLI group consisted of students whose scores fell within the range of impairment on both the language and nonverbal intelligence measures.

When the students were tested in eighth grade, the NL group consisted of 316 students: 178 boys and 138 girls. The SLI group was composed of 59 students, 38 boys and 21 girls, with a mean age of 13.93 years (*SD* = 0.40). The LC group included 72 students, 38 boys and 34 girls, with a mean age of 13.95 years (*SD* = 0.38). The NLI group contained 80 students, 39 boys and 41 girls, with a mean age of 13.92 years (*SD* = 0.37). Table 1 lists descriptive statistics for the measures of nonverbal cognition and language by diagnostic group.

Procedure

Discourse processing—Three narrative passages were adapted from Crais and Chapman (1987) and Kertoy and Goetz (1995) and presented to participants via headphones and high-quality audio recorders. The narratives were read by a female speaker with Northern Midland dialect at a normal speaking rate and presented at a constant conversational volume. The Appendix contains the stories and associated questions. The stories centered on topics derived from international fables. Central characters were animals or humans facing challenges. The stories presented background information needed to comprehend the story, then introduced the characters and events. The stories ranged from 207 to 268 words in length.

The students were tested individually. An examiner provided the following instructions:

> Now I'm going to play you some short stories on the tape recorder. After each story, I'll ask you some questions. I'm going to tape record your answers and write them down on my sheet. Do you have any questions? Okay, let's listen.

The student then listened to an audio recording of the first story using headphones. After the entire story had been presented, the examiner stopped the tape and asked the student eight questions: four premise questions and four inference questions. Two of the inference questions, called *adjacent inferences*, were based on information contained in the same sentence or in an adjacent sentence. The remaining two inference questions, termed *distant inferences*, were based on information separated by four or more sentences. The four premise questions were based on information specifically stated in the story. Two of the premise questions—*adjacent premises*—provided the basis for the adjacent inference questions, and two of the premise questions—*distant premises*—provided the basis for the distant inference questions. This design enabled us to assess whether or not the difference between subgroups' answering of inference questions was related to difference in retention or comprehension of information that was specifically stated in the story. The procedure was then repeated for the remaining two stories. The order in which the stories were presented was counterbalanced across the three stories. The examiner asked the same questions in the same order to all participants for each story. The order of the questions was determined using a semirandom approach in which related premise and inferences questions were not

presented consecutively, and no more than three premise or three inference questions were presented in a row.

Working memory—A battery of working memory tasks was administered, and the relationship between performance on these tasks and the discourse processing task was assessed. Table 2 reports the descriptive statistics for the working memory tasks by eighth-grade diagnosis and indicates significant pairwise contrasts between the groups. This battery consisted of four verbal working memory measures and a spatial working memory measure. The nonword repetition task (NRT; Dollaghan & Campbell, 1998) assessed phonological memory. Students repeated 16 nonsense words that ranged from one to four syllables in length, began and ended with consonants, and contained only tense vowels. The Auditory Working Memory (AWM) subtest of the Woodcock-Johnson–3 (Woodcock, McGrew, & Mather, 2001) measured students' ability to categorize while retaining sequences. Following the auditory presentation of a series of digits and words, students repeated the words in sequential order and then repeated the digits in the order in which they were presented. The CLPT (Gaulin & Campbell, 1994) assessed working memory while comprehending and producing language. Students stated whether spoken sentences were true or false, and after a two- to six-sentence set had been presented, they were asked to recall the final word in each sentence in the set. The grammatical judgment listening span task (GJLST; Ellis Weismer, 2006) assessed verbal working memory during linguistic processing. Students stated whether spoken sentences were grammatically acceptable, and after a set of two to six sentences had been presented, they were asked to recall the final word in each sentence in the set. The spatial working memory task (SWMT; Ellis Weismer, 2006), which was adapted from Russell, Jarrold, and Henry (1996), evaluated nonverbal working memory using a complex odd-one-out paradigm. Students were presented with two to six sets of three complex shapes on a computer screen and asked to point to the shape that did not match the other two. After all of the sets had been presented at a level (i.e., two sets, three sets, etc.), students pointed to the locations of the shape that did not match the others on a blank grid on the computer screen.

Results

Distance

Our first research question asked whether or not inference questions based on distant information would be more difficult to answer than inference questions based on adjacent information. The mean number of adjacent inference questions answered correctly was 4.94, with a standard deviation of 1.20. The mean number of distant questions answered correctly was 3.51, with a standard deviation of 1.44. A repeated measures analysis of variance (ANOVA) revealed a significant effect of distance, $F(1, 526) = 509.53$, $\eta_p^2=.49$, $p = .000$, indicating that distant inference questions were significantly more difficult to answer than adjacent inference questions. The effect size was large, suggesting a robust effect for the distance between premises upon inference construction.

Diagnostic Group

Our second research question pertained to whether there were group differences in the accuracy of response to inference questions. Table 3 summarizes the descriptive statistics for the discourse processing task broken down by current diagnosis. In order to correctly answer the inference questions, the adolescents would need to understand the premises on which the inferences were based. Due to the counts nature of the outcome variable (i.e., number of correct item responses), the analysis was performed assuming a binomial distribution for the outcome with a logit link to dummy-coded group predictors and a covariate (premise accuracy). The model consequently characterizes the effects of group and covariate on the

probability of a correct item response. The model was fit using the generalized linear model procedure implemented in SPSS. In each case, an omnibus Wald test was applied to determine the significance of group effects, followed by pairwise comparison through Fisher's least significant difference (LSD) procedure to evaluate specific group effects. The Wald test revealed significant effects for the premise covariate for adjacent inferences, $\chi^2(1, N = 527) = 147.69$, $p = .000$, and for distant inferences, $\chi^2(1, N = 527) = 45.71$, $p = .000$. The remaining group effects were also significant for both adjacent inferences, $\chi^2(3, N = 527) = 23.33$, $p = .000$, and distant inferences, $\chi^2(3, N = 527) = 61.26$, $p = .000$.

Pairwise comparisons using LSD were performed to assess the differences between the diagnostic groups on adjacent and distant inference questions. For adjacent inference questions, the NL group performed significantly better than both the SLI group ($p = .031$) and the NLI group ($p = .000$), and the LC group significantly outperformed the NLI group ($p = .020$). There was not a significant difference between the NL and LC groups on adjacent inferences ($p = .130$), or between the SLI and LC groups ($p = .558$) or SLI and NLI groups ($p = .071$). On distant inference questions, the NL group performed significantly better than each of the other groups ($p = .000$). Adolescents in the LC group performed significantly better than those in the NLI group ($p = .008$). The LC and SLI groups did not significantly differ on distant inference accuracy ($p = .352$). There was also no significant difference between the SLI and NLI groups on distant inference accuracy ($p = .111$).

Error Patterns

The third research question asked whether the diagnostic groups differed in the types of errors they displayed on the discourse processing task. The same types of errors were observed in each of the groups. Errors were called *omissions* when the student did not give a response or stated that he or she did not know the answer. *Incomplete* responses were those for which the student gave an answer that was partially correct. For example, responding to the distant inference question "How was the girl able to get Willy away from the alligator" in "The Girl and the Alligator" story (see the Appendix) with "by touching the alligator's stomach," without adding that rubbing the alligator's stomach made him fall asleep, was coded as an incomplete response. The third type of error consisted of answers that were simply incorrect. For the different diagnostic groups, we computed the proportion of each error type for adjacent inferences and for distant inferences by dividing the number of errors of that type (omission, incomplete, or incorrect answers) into the total number of errors. We conducted an analysis to examine the effects of diagnosis and distance (adjacent or distant) on the types of errors observed on inference questions, with each type of error analyzed separately.

ANOVA findings indicated a main effect of diagnostic group for omissions, $F(3, 513) = 9.29$, $\eta_p^2 = .05$, $p = .000$, and incorrect answers, $F(3, 509) = 4.19$, $\eta_p^2 = .02$, $p = .006$, but not for incomplete answers. We performed pairwise comparisons using LSD to assess differences between groups. The SLI group produced a significantly higher proportion of omission errors than the LC group ($p = .002$) and the NL group ($p = .001$) and did not differ significantly from the NLI group ($p = .847$). The NLI group produced a significantly higher proportion of omissions than the LC group ($p = .000$) and the NL group ($p = .000$), who did not differ from one another. Pairwise comparisons revealed that only the NL and NLI groups were significantly different from each other with respect to the proportion of incorrect answers ($p = .001$). A significant main effect of distance was found for incorrect answers, $F(1, 509) = 18.26$, $\eta_p^2 = .04$, $p = .000$, and incomplete answers, $F(1, 513) = 50.92$, $\eta_p^2 = .09$, $p = .000$, such that more errors occurred on distant inference questions than on adjacent inferences. There was no significant interaction between diagnostic group and distance on

any of the error types. It should be noted that when significant differences were found in the error analyses, the effect sizes were small, indicating that these differences were not robust.

Working Memory

Our fourth research question focused on the role of working memory in predicting performance on distant inference items. Pearson correlation coefficients revealed that each of the working memory measures was significantly correlated with distant inference accuracy (see Table 4). We performed a series of multiple regression analyses to assess the contribution of language, cognitive, and working memory measures to distant inference question accuracy. First, we entered the various language measures as predictors into a hierarchical linear regression model, starting with the grammatical measures (CELF Following Directions, CELF Sentence Repetition) and followed by the vocabulary measures (PPVT–R, CREVT). When the four language measures were used to predict distant inference performance, the model accounted for 22% of the variance, $F(4, 526) = 38.67$, $p = .000$. As indicated in Table 5, each of the language measures except the CREVT had a significant standardized beta.

Next, a hierarchical regression model was fit to the data for the working memory measures. The verbal working memory measures were entered first, followed by the spatial working memory task (AWM, CLPT, NRT, GJLST, and SWMT). The working memory measures predicted 19% of the variance in distant inference accuracy, $F(5, 524) = 24.86$, $p = .000$, with the CLPT and SWMT emerging as significant predictors, as summarized in Table 6.

In the final model, we used stepwise regression in which we entered the three significant language predictors and the two significant working memory measures as predictors, along with PIQ. A three-step regression model significantly predicted distant inference accuracy, accounting for just over 25% of the variance, $F(3, 526) = 59.19$, $p = .000$. As seen in Table 7, two language measures and a verbal working memory measure emerged as significant predictors. The PPVT–R accounted for 19% of the variance, the CLPT accounted for an additional 6% of significant unique variance, and the CELF Following Directions subtest accounted for less than 1% additional significant variance.

Discussion

Distance

Inference questions that were based on widely spaced information were more difficult to answer than inference questions based on adjacent information. This robust effect demonstrated that adolescents with normal language, low cognition, SLI, and non-SLI have difficulty linking together widely spaced information in order to generate inferences, as do children and adolescents with poor reading comprehension (Catts et al., 2006; Yuill et al., 1989) or traumatic brain injury (Moran & Gillon, 2005), and even typical adults (Lea et al., 2005).

Diagnostic Group

Adolescents with normal language and cognition were significantly more likely to answer inference questions accurately than adolescents with language impairments, extending the findings of studies of children with SLI (Bishop & Adams, 1992; Ellis Weismer, 1985) to a large population-based sample of adolescents. The performance of the adolescents with SLI did not significantly differ from that of adolescents with low cognition or NLI. Humphries et al. (2004) and Worling et al. (1999) also found that children with nonverbal learning disabilities performed similarly on inferencing tasks to children with verbal impairments. The present study indicates that these similarities persist when it is necessary to link more

widely spaced information. The present study further adds that adolescents with SLI do not significantly differ from those with NLI on inferencing, although those with normal language and cognition and those with low cognition but intact language do outperform those with NLI.

According to Kintsch's construction-integration model, the understanding of linguistic input, general world knowledge, and working memory are all necessary for the construction of inferences. In the present study, we used PIQ—scores on the Block Design and Picture Completion subtests of the WISC–III—as a gross index of "general world knowledge." These subtests provide information about reasoning and problem-solving ability but do not provide specific content about world knowledge. However, measures of world knowledge that include language, such as vocabulary tests, would underestimate the world knowledge of individuals with language impairment, rendering such measures inappropriate for use in this investigation. It was reasoned that adolescents who demonstrated deficits in reasoning and problem solving would be likely to present with decreased world knowledge as compared with adolescents who exhibited normal range PIQ; however, further research is needed to evaluate the use of PIQ as a measure of general world knowledge.

The adolescents in the NL group adequately performed tasks involving language comprehension, world knowledge, and working memory, and it is therefore not surprising that they outperformed each of the other groups on distant inference accuracy. The LC group of adolescents' linguistic abilities were intact, and their working memory was significantly better than that of groups with language impairment, but they have decreased general world knowledge compared with the NL and SLI groups. The SLI group demonstrated typically developing general world knowledge, but they displayed lower linguistic and working memory abilities than the LC and NL groups. The similar performance of the LC and SLI groups could indicate that deficits in either general knowledge or linguistic abilities and verbal working memory can result in difficulty with constructing inferences, particularly when the information needed to generate the inference is widely spaced. The adolescents with NLI exhibited deficiencies in all three of the areas needed for inference construction. Their performance on distant inferences was significantly lower than that of the two groups without language impairment but did not differ significantly from that of the SLI group. These findings suggest either that PIQ was not a reasonable proxy for general world knowledge or that the impact of world knowledge in constructing inferences is less heavily weighted than the factors of linguistic comprehension and working memory.

Error Patterns

Error analyses revealed a significant main effect of group for omission errors. Students with language impairment produced a higher proportion of omissions than did those with typical language. This is not surprising given that adolescents with language impairment have deficits in language production as well as comprehension and may have had difficulty formulating responses to the questions. There was also a main effect of group for wrong answers, with only the NL and NLI groups demonstrating significant differences. Students with NLI are deficient in general world knowledge, verbal working memory, and comprehension of linguistic input, whereas students with SLI or LC exhibit deficits in only one or two of these areas, rather than all three. The students with SLI or LC may have been able to use their areas of strength to compensate somewhat for their deficits, resulting in making fewer outright errors than the students with NLI. There was a significant main effect of distance for wrong answers and incomplete answers, indicating that questions requiring students to link widely spaced information resulted in more errors that were simply incorrect or only partially correct than questions that were based on adjacent information. The lack of a significant main effect of distance for omission errors may be due to the students with poor language production (who were most likely to produce this type of error) having difficulty

generating responses regardless of their ability to link the information across the passage to make the inference.

Working Memory

Regression analyses indicated that when working memory measures were considered independently, both a verbal working memory measure and a spatial working memory measure were significant predictors of distant inference construction. This finding supports the assertion that working memory plays an important role in inference construction as posited by Kintch's model. When considered in combination with other factors, one of the verbal working memory measures, the CLPT, significantly contributed a unique, albeit small, amount of variance to distant inference performance beyond that explained by general language or cognitive abilities. It is important to acknowledge, however, that the impact of variation in working memory on verbal inference construction was relatively modest for this particular discourse processing task.

Spatial working memory did predict unique variance in inference construction, providing some evidence that domain-general working memory may be involved in inferencing; however, spatial working memory was no longer a significant predictor when language and PIQ measures were entered into the model. The CLPT, a measure of verbal working memory, remained a significant predictor of inference construction even when language and PIQ measures were included, providing evidence for the impact of domain-specific verbal working memory on inferencing. Further research is necessary to assess the contributions of domain-general versus domain-specific working memory to the construction of inferences.

Summary and Implications

Eighth graders answered inference questions based on adjacent information more accurately than they did inference questions based on distant information. When accuracy of answers to premise questions was controlled, adolescents with SLI performed significantly worse than those with normal language and cognition, and no significant differences were found between the adolescents with SLI and those with low cognition or NLI. These results indicate that adolescents who are deficient in language comprehension, verbal working memory skills, and/or general world knowledge—all of which are needed for inference construction (Kintsch, 1988)—have difficulty making inferences, especially when it is necessary to hold information active in working memory over a distance while processing new information. Verbal working memory played a role in the ability to accurately answer distant inference questions for all of the diagnostic groups. Further research is needed to fully assess the complex relationships among comprehension of linguistic input, general world knowledge, and verbal working memory in the construction of inferences.

Acknowledgments

Appendix. Stories and associated questions

The Girl and the Alligator

Everyone who lives around alligators knows that if you rub an alligator's stomach, it will soon go to sleep. A long time ago, a little girl lived in a far-away and exotic place. It was called "Danger Jungle" because there were so many alligators hiding in the bushes and swamps. One day, a vicious alligator came to the little girl's house and grabbed her dog "Willy." The alligator took Willy to the bank of the river and laid down in the sun. The alligator held Willy down with its enormous, strong tail. While he held Willy, the alligator rolled onto its back and began to sharpen its teeth so it could feast on the little dog. Meanwhile, the little girl had been helping her mother with chores in the house. She thought she should check on Willy. When she saw that Willy was gone, she knew she had to find him quickly. She had an idea. She stood very still and listened for Willy's barking. Then she ran as fast as she could until she found him and the alligator. She was breathing hard because she had run a long way. As the girl watched the alligator lie on its back, she remembered what people had told her about alligators. She slowly crept up to the alligator close enough to touch him, and soon he was fast asleep. She hoped Willy would stay quiet even though he was scared. The little girl pulled Willy from under the alligator and ran home as fast as her feet would carry her.

Adjacent Premise Questions

- What did the little girl hope that Willy would do when she pulled him from under the alligator?
- What did the girl do right before she ran to the river to get Willy?

Distant Premise Questions

- What happens if you rub an alligator's stomach?
- What was the girl doing before she realized her dog was gone?

Adjacent Inference Questions

- Why did the little girl hope that Willy would stay quiet?
- How did the girl know where to look for Willy?

Distant Inference Questions

- How do you know that the little girl's house was far from the river?
- How was the little girl able to get Willy away from the alligator?

The Donkey and the Wolf

Donkeys can kick extremely hard, so it's important to watch out for their kicking feet. One day just before dinner, a very thin wolf approached a pasture where a donkey was grazing. He was so excited to see the donkey that he ran out into the pasture as fast as he could. To his surprise, the donkey had disappeared. The wolf thought his mind was playing tricks on him, but really the donkey had gone behind a boulder. When the donkey finally appeared from behind the boulder, he was limping and pretending to have hurt his foot. The wolf said, "You can't get away from me, Donkey. You will soon be my dinner!" The donkey said, "Oh please, Mr. Wolf, don't eat me!" Then he said, "But if you do eat me, first you'd better get the thorn out of my foot or it may stick in your throat when you swallow me." The wolf thought the donkey had a point, so he lifted up the donkey's foot to look for the thorn. The donkey laughed to himself later to think that he had tricked the wolf and gotten away.

Adjacent Premise Questions

- What did the wolf look like?

- What part of a donkey should you be careful to stay away from?

Distant Premise Questions

- What did the donkey pretend was wrong with him?

- What happened when the wolf first ran out into the pasture?

Adjacent Inference Questions

- How do you know the wolf hadn't eaten in a long time?

- What can donkeys do to hurt you?

Distant Inference Questions

- Why did the donkey hide behind a boulder when he first saw the wolf?

- How did the donkey get away from the wolf?

The King's Ducks

Once upon a time, there was a king who lived in a gorgeous castle on the edge of a beautiful lake. This king had two special ducks that he was very fond of. He told his young son that if he could determine which duck was the male duck, he would give the son all his riches. The son thought that surely the bigger duck was the male duck, but he figured he had better do some more investigating before making a decision. The son traveled out to the countryside and requested help from a man he saw walking on a country road. The farmer was amazed to see the king's son outside of the castle grounds. "Would you be so kind as to help me, sir?" said the boy. "I need to know how to tell male ducks from female ducks. I would appreciate any information you can give me, since my future depends on it." The farmer told the son to take the ducks out of the water and set them on the shore. Then, he told him to watch carefully to see which duck would enter the water first. The farmer said that the first to enter the water was always the male duck. The boy returned to the castle to try out the farmer's advice. He was glad he had listened to the farmer, because he was surprised to discover which duck was the male. The following day, he informed the king that he knew which duck was the male duck. The king was very pleased and kept his promise to his son.

Adjacent Premise Questions

- Which duck did the farmer say would enter the water first?

- Where did the young son find someone to help him?

Distant Premise Questions

- What would the king give his son if he could figure out which duck was the male duck?

- Before he talked to the farmer, which duck did the son think was the male?

Adjacent Inference Questions

- Who was the man that was walking on the country road?

- How did the young son determine which of the King's ducks was male?

Distant Inference Questions

- How do you know the son got the king's riches?

- Why was the son surprised when he figured out which duck was the male duck?

References

Bishop DVM, Adams C. Comprehension problems in children with specific language impairment: Literal and inferential meaning. Journal of Speech and Hearing Research. 1992; 35:119–129. [PubMed: 1735960]

Cain K, Oakhill JV, Bryant PE. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. Journal of Educational Psychology. 2001; 96:31–42.

Catts HW, Adolf SM, Ellis Weismer S. Language deficits in poor comprehenders: A case for the simple view of reading. Journal of Speech, Language, and Hearing Research. 2006; 49:278–293.

Crais ER, Chapman RS. Story recall and inferencing skills in language/learning-disabled and non-disabled children. Journal of Speech and Hearing Disorders. 1987; 52:50–55. [PubMed: 3807345]

Dollaghan CA, Campbell TF. Nonword repetition and child language impairment. Journal of Speech and Hearing Research. 1998; 52:1136–1146.

Dunn, LM.; Dunn, LM. Peabody Picture Vocabulary Test—Revised. Circle Pines, MN: AGS; 1981.

Ellis Weismer S. Constructive comprehension abilities exhibited by language-disordered children. Journal of Speech and Hearing Research. 1985; 28:175–184. [PubMed: 4010247]

Ellis Weismer, S. Examining the role of processing limitations in specific language impairment. Keynote lecture presented at the Symposium on Research in Child Language Disorders; Madison, WI. 2006 June.

Ellis Weismer S, Evans J, Hesketh LJ. An examination of verbal working memory capacity in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 1999; 42:1249–1260.

Ellis Weismer S, Plante E, Jones M, Tomblin JB. A functional magnetic resonance imaging investigation of verbal working memory in adolescents with specific language impairment. Journal of Speech, Language, and Hearing Research. 2005; 48:405–425.

Fey M, Catts H, Proctor-Willams K, Tomblin JB, Zhang X. Oral and written story composition skills of children with language impairment. Journal of Speech, Language, and Hearing Research. 2004; 47:1301–1318.

Gaulin C, Campbell T. Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. Perceptual and Motor Skills. 1994; 79:55–64. [PubMed: 7991333]

Humphries T, Cardy JO, Worling DE, Peets K. Narrative comprehension and retelling abilities of children with nonverbal learning disabilities. Brain and Cognition. 2004; 56:77–88. [PubMed: 15380878]

Just MA, Carpenter PA. A capacity theory of comprehension: Individual differences in working memory. Psychological Review. 1992; 99:122–149. [PubMed: 1546114]

Kertoy MK, Goetz KM. The relationship between listening performance on the sentence verification technique and other measures of listening comprehension. Contemporary Educational Psychology. 1995; 20:320–339.

Kintsch W. The role of knowledge in discourse comprehension: A construction integration model. Psychological Review. 1988; 95:163–182. [PubMed: 3375398]

Lea RB, Mulligan EJ, Walton JL. Accessing distant premise information: How memory feeds reasoning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:387–395.

Leslie, L.; Caldwell, J. Qualitative Reading Inventory—3. Boston, MA: Allyn & Bacon; 2001.

MacDonald MC, Christiansen MH. Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review. 2002; 109:35–54. [PubMed: 11863041]

Mainela-Arnold E, Evans J. Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. Journal of Speech, Language, and Hearing Research. 2005; 48:897–909.

Marton K, Schwartz RG. Working memory capacity and language processes in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2003; 46:1138–1153.

Montgomery J. Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? American Journal of Speech-Language Pathology. 2002; 11:77–91.

Montgomery J, Evans J. Complex sentence comprehension and working memory in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2009; 52:269–288.

Moran C, Gillon G. Inference comprehension of adolescents with traumatic brain injury: A working memory hypothesis. Brain Injury. 2005; 19:743–751. [PubMed: 16175835]

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 and Hearing Research. 2004; 47:199–211.

Russell J, Jarrold C, Henry L. Working memory in children with autism and moderate learning disabilities. Journal of Child Psychology and Psychiatry and Allied Disciplines. 1996; 37:673–686.

Semel, E.; Wiig, E.; Secord, W. Clinical Evaluation of Language Fundamentals—Third Edition. San Antonio, TX: Psychological Corporation; 1995.

Snyder LS, Downey DM. The language-reading relationship in normal and reading-disabled children. Journal of Speech and Hearing Research. 1991; 34:129–140. [PubMed: 2008067]

Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. Journal of Speech, Language, and Hearing Research. 1997; 40:1245–1260.

Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. Journal of Speech, Language, and Hearing Research. 1996; 39:1284–1294.

Virtue S, Haberman J, Clancy Z, Parrish T, Beeman M. Neural activity of inferences during story comprehension. Brain Research. 2006; 1084:104–114. [PubMed: 16574079]

Virtue S, van den Broek P. Hemispheric processing of anaphoric inferences: The activation of multiple antecedents. Brain and Language. 2004; 93:327–337. [PubMed: 15862857]

Virtue S, van den Broek P, Linderholm T. Hemispheric processing of inferences: The effects of textual constraint and working memory capacity. Memory & Cognition. 2006; 34:1341–1355.

Wallace, G.; Hammill, D. Comprehensive Receptive and Expressive Vocabulary Test. Austin, TX: Pro-Ed; 1994.

Wechsler, D. Wechsler Intelligence Scales for Children—III. San Antonio, TX: Psychological Corporation; 1991.

Woodcock, R.; McGrew, K.; Mather, N. Woodcock–Johnson Psychoeducational Battery—III. Itasca, IL: Riverside; 2001.

Worling DE, Humphries T, Tannock R. Spatial and emotional aspects of language inferencing in nonverbal learning disabilities. Brain and Language. 1999; 70:220–239. [PubMed: 10550228]

Yuill N, Oakhill J, Parkin A. Working memory, comprehension ability and the resolution of text anomaly. British Journal of Psychology. 1989; 80:351–361. [PubMed: 2790393]

Table 1

Descriptive statistics for nonverbal cognitive ability and language measures by eighth-grade diagnosis

Measure	NL *M* (*SD*)	LC *M* (*SD*)	SLI *M* (*SD*)	NLI *M* (*SD*)
WISC PIQ	104.36 (11.60)	79.04 (7.40)	97.83 (9.42)	73.49 (10.51)
PPVT–R	101.24 (13.59)	91.53 (12.06)	82.90 (8.96)	76.26 (9.40)
CREVT	96.99 (11.65)	90.26 (9.68)	82.44 (8.19)	81.71 (7.14)
CELF: FD	9.33 (2.84)	8.36 (2.34)	5.44 (1.98)	4.28 (1.54)
CELF: SR	8.98 (2.51)	8.04 (2.15)	4.54 (1.45)	4.64 (1.82)
CELF: *z* score	−0.57 (0.79)	−0.81 (0.62)	−1.95 (0.47)	−2.02 (0.48)
Language composite	0.08 (0.79)	−0.49 (0.62)	−1.50 (0.40)	−1.89 (0.70)

Note. NL = normal language; LC = low cognition; SLI = specific language impairment; NLI = nonspecific language impairment; WISC PIQ = Wechsler Intelligence Scales for Children, Performance IQ scale; PPVT–R = Peabody Picture Vocabulary Test—Revised; CREVT = Comprehensive Receptive and Expressive Vocabulary Test; CELF: FD = Clinical Evaluation of Language Fundamentals: Following Directions subtest; SR = sentence repetition.

Table 2

Descriptive statistics and significant pairwise contrasts for working memory measures by eighth-grade diagnosis

Measure	NL$_a$ M (SD)	LC$_b$ M (SD)	SLI$_c$ M (SD)	NLI$_d$ M (SD)
CLPT	31.37 (5.19)$_{c,d}$	30.18 (4.84)$_{c,d}$	25.15 (5.15)$_{a,b}$	24.30 (4.99)$_{a,b}$
GJLSPT	23.83 (5.01)$_{b,c,d}$	22.05 (4.89)$_{a,c,d}$	17.58 (4.70)$_{a,b}$	16.84 (5.01)$_{a,b}$
NRT	90.96 (6.65)$_{c,d}$	89.69 (7.09)$_{c,d}$	84.79 (9.49)$_{a,b}$	83.21 (8.33)$_{a,b}$
SWMT	67.51 (8.56)$_{b,c,d}$	61.43 (10.78)$_{a,d}$	58.75 (9.91)$_{a,d}$	52.01 (12.45)$_{a,b,c}$
AWM	106.58 (11.78)$_{b,c,d}$	101.11 (12.46)$_{a,c,d}$	91.86 (13.61)$_{a,b}$	88.33 (11.35)$_{a,b}$

Note. CLPT = competing language processing task; GJLSPT = grammatical judgment listening span task; NRT = nonword repetition task; SWMT = spatial working memory task; AWM = Auditory Working Memory subtest of the Woodcock-Johnson–3. Subscripts indicate significant pairwise differences ($p < .01$) between the group within that column and the other groups indicated.

Table 3

Descriptive statistics for discourse processing task by eighth-grade diagnosis

Measure	NL_a M (SD)	LC_b M (SD)	SLI_c M (SD)	NLI_d M (SD)
Adjacent premise	4.84 (1.02)	4.69 (1.08)	3.93 (1.17)	3.37 (1.45)
Adjacent inference	5.26 (0.89)$_{c,d}$	5.02 (1.05)$_c$	4.56 (1.19)$_a$	3.86 (1.67)$_{a,b}$
Distant premise	5.26 (0.92)	4.99 (1.19)	4.44 (1.39)	3.95 (1.52)
Distant inference	3.96 (1.27)$_{b,c,d}$	3.28 (1.40)$_{a,d}$	2.90 (1.26)$_a$	2.41 (1.43)$_{a,b}$

Note. The range of possible scores is 0–6. Subscripts indicate significant pairwise differences ($p < .05$) between the group within that column and the other groups indicated.

Table 4

Correlations between working memory measures and distant inference accuracy

Working memory measure	Distant inference accuracy
Competing language processing task	.41[**]
Grammatical judgment listening span task	.33[**]
Nonword repetition task	.28[**]
Spatial working memory task	.34[**]
Auditory Working Memory subtest of Woodcock-Johnson–3	.33[**]

[**]
$p < .01$.

Table 5

Regression analysis of language measures as predictors of distant inference accuracy

Predictor	*B*	*SE B*	β	*p*
CELF-FD	0.071	0.022	.159	.001[**]
CELF-SR	0.063	0.026	.130	.016[*]
PPVT	0.027	0.005	.301	.000[*]
CREVT	−0.004	0.006	−.036	.491

Note. $R^2 = .23$. $p = .000$.

[*] $p < .05$.

[**] $p \leq .001$.

Table 6

Regression analysis of working memory measures as predictors of distant inference accuracy

Predictor	*B*	*SE B*	β	*p*
AWM	0.008	0.005	.083	.120
CLPT	0.061	0.015	.250	.000**
NRT	0.013	0.008	.073	.114
GJLST	0.004	0.015	.016	.789
SWMT	0.015	0.007	.116	.024*

Note. AWM = Auditory Working Memory subtest of the Woodcock-Johnson–3. R^2 = .19. *p* = .000.

*
p < .05.

**
p < .001.

Table 7

Stepwise regression analysis of cognitive, language, and working memory measures as predictors of distant inference accuracy

Step	Predictor	B	SE B	β	p	R^2	ΔR^2
1	PPVT	0.040	0.004	.434	.000	.188	
2	PPVT	0.029	0.004	.317	.000		
	CLPT	0.065	0.010	.267	.000	.246	.058**
3	PPVT	0.026	0.004	.282	.000		
	CLPT	0.054	0.011	.223	.000		
	CELF-FD	0.050	0.021	.111	.019	.253	.008*

*
$p < .05$.

**
$p < .001$.