Understanding Narrative Texts: Inferences Made During Reading by Children With and Without Language Impairments

Brenna McClintock

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Child Study) at

Concordia University

Montreal, Quebec, Canada

December, 2011

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Brenna McClintock

Entitled: Understanding Narrative Texts: Inferences Made During Reading by Children With and Without Language Impairments

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Child Study)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Holly Recchia Chair

Sandra Martin-Chang Examiner

Nina Howe Examiner

Diane Pesco Supervisor

Approved by _____
Chair of Department or Graduate Program Director

_____ 2012 _____
Dean of Faculty

Abstract

Understanding Narrative Texts: Inferences Made During Reading by Children With and Without Language Impairments

Brenna McClintock

Successful reading comprehension relies on a number of abilities, including decoding and using linguistic and world knowledge to link text propositions. While difficulties with reading comprehension are well documented for children with Specific Language Impairment (SLI), only a handful of studies have investigated the inferences children with SLI make as they read. The present study examines the inferences made by children with SLI during and after reading, with particular attention to causal inferences, considered critical to good comprehension. The common procedure for assessing inferences - asking comprehension questions after reading - was combined with the more novel approach of eliciting children's thoughts about the text as they read ("think-aloud").

Twelve children with SLI (*M* age = 10;2) were compared to 12 age-matched peers with typical language development (TLD). The SLI group had lower scores than the TLD group on literal comprehension questions, and inferential questions requiring informational and causal inferences. The think-aloud data were consistent with these findings, and additionally showed that the causal inferences of the SLI group were less accurate than in the TLD group. Nevertheless, both groups had higher scores on comprehension questions after engaging in the think-aloud, with the improvement in the SLI group driven by a significant improvement in causal comprehension. The results are examined in light of children's performance on measures of oral language and working

memory, and discussed in terms of reading comprehension theories. Practical applications of the think-aloud procedure are also considered.

Acknowledgments

I would like to acknowledge the support, wisdom, and guidance that my thesis supervisor, Dr. Diane Pesco, has provided to me so graciously throughout the entire thesis process. Her patience and sense of humour have made this project possible. I would also like to express my gratitude to my committee members, Dr. Nina Howe and Dr. Sandra Martin-Chang, whose insight and advice have been invaluable.

I would like to thank my friends and colleagues who so eagerly found their inner child to participate in the piloting phase of the study.

I would like to thank the teachers, who so willingly collaborated on this project and were a pleasure to work with. Finally, I would like to offer my most sincere thanks to the students who participated in this study for their time, participation, and wonderful thoughts.

Table of Contents

List of Figures

List of Tables

Introduction

In the elementary school years, children progress from "learning to read" to "reading to learn". Reading to learn, a major component of academic success, is a skill that requires good reading comprehension (Chall, 1983); many children, however, struggle with fully comprehending what they have read. In 2010, approximately 15 to 20% of children in the U.S. were considered to have a reading disability (National Institute of Child Health and Human Development [NICHHD], 2010). Literacy levels in Canada are also of concern. Approximately 30% of 15-year-old Canadians have not yet achieved a level of literacy that is considered to be a functional level in a knowledge economy (Statistics Canada, 2010).

Difficulties with reading comprehension have been attributed to a number of different factors. These include difficulties with decoding (Gough & Tunmer, 1986) and reading fluency (Cutting & Scarborough, 2006; Kim, Petscher, Schatschneider, & Foorman, 2010), weaknesses in oral language comprehension (Catts, Fey, Zhang, & Tomblin, 1999; Clarke, Snowling, Truelove, & Hulme, 2010), and difficulties with cognitive abilities such as working memory (Alloway & Alloway, 2010; Cain, Oakhill, & Lemmon, 2004; Perfetti & Zhang, 1996) and attention (McInnes, Humphries, Hogg-Johnson, & Tannock, 2003). Children's failure to actively engage with text by asking questions, monitoring their comprehension, and inferencing also contributes to poor reading comprehension (Shanahan, 2006).

Inferencing is broadly defined as a form of reasoning used to relate explicit and implied information (Colman, 2009; NICHHD, 2006). In the literature on reading, the term is defined as "the abstraction of information that is not explicitly presented"

(Botting & Adams, 2005, p. 50), and includes the reader's ability to draw conclusions by connecting propositions in a text ("text-to-text" inferences) and integrate textual information with world knowledge ("text-to-world" inferences) (van Kleeck, 2008; Zwiers, 2010). For children, skill with both kinds of inferences is associated with good comprehension (Bowyer-Crane & Snowling, 2005). Children who read below age and grade level, including many with Specific Language Impairment (SLI), have particular difficulty with making text-to-world inferences (Cain & Oakhill, 1999; Norbury & Bishop, 2002). Given that children with SLI are a major focus of the present study, the nature of SLI will be discussed at length below. In general terms, SLI refers to impaired first language development in the absence of other major developmental delays.

Models of Reading and Reading Impairment

The Simple View of Reading

The simple view of reading is an influential model of reading proposed by Hoover and Gough (1990). In this model, reading is the product of decoding ability and linguistic comprehension (R = D x L). Decoding is the ability to identify printed words efficiently, whereas linguistic comprehension is the ability to interpret lexical information in order to derive meaning from information at the sentence-level in text, and in conversation. Decoding and linguistic comprehension are equally important; neither component alone is sufficient for successful reading. Reading occurs successfully when printed words are identified accurately, and when meaning is successfully derived from words and the syntactic relationships between them (Cutting & Scarborough, 2006). Amongst good readers, decoding and linguistic comprehension are positively related: improvement in one component leads to an improvement in the other. As the two

components improve, so does reading comprehension (Hoover & Gough, 1990).

Support for the simple view of reading comes from studies in which the contribution of decoding and oral language skills to reading comprehension was investigated. The contribution made by the combination of the two components varies significantly according to the measures used, but is consistently large: 47% in Kendeou, van den Broek, White, and Lynch (2009), and as high as 72% in Cutting and Scarborough (2006). Furthermore, the weight of each component changes as children become proficient readers. Kendeou et al. (2009) found that amongst typically developing preschoolers, oral language skills predicted as much as 28% of the variance in decoding skill. As children learned to decode more fluently, oral language played less of a role in decoding ability, and oral language and decoding began to predict reading comprehension independently. By second grade, both decoding and oral language skills made significant and independent contributions to reading comprehension, with oral language predicting more of the variance in reading comprehension than decoding.

Catts et al. (1999) found that scores on oral language measures of kindergarten-aged children were the single, strongest predictors of reading comprehension and word recognition difficulties in grade 2. Fifty-seven percent of the poor readers had been identified in kindergarten as having impairments in receptive language, while 50% had been identified as having expressive language problems (implying that some children had mixed expressive and receptive difficulties). These percentages were four to five times greater than those found among good readers. Thus, both decoding and oral language skills play an important role in a child's ability to comprehend what is read. As children learn to skillfully decode text, the importance of oral language skills to reading

comprehension increases. This finding has important implications for students who experience weaknesses in either, or both, decoding ability and oral language.

The Simple View and Reading Impairments

For good readers, the simple view states that the relationship between decoding and linguistic comprehension is positively related. This, however, is often not the case for children with reading disabilities, who display three distinct reading profiles (Gough & Tunmer, 1986). In two of these profiles, decoding and linguistic comprehension are dissociated, resulting in (1) classic dyslexia or (2) hyperlexia (more recently referred to in the literature as poor comprehending). In dyslexia, decoding is impaired while linguistic comprehension is average. In poor comprehenders, decoding ability is at or above levels expected for chronological age, while linguistic comprehension is impaired (Clarke et al., 2010). The third profile, referred to as 'garden variety reading disability' or 'generally poor readers', involves deficiencies in both decoding and linguistic comprehension. For this group, performance on both components falls below what would be expected for chronological age.

Recent research has shown that children with SLI are a heterogeneous group. Some fit the poor comprehender profile while others fit the generally poor reader profile (Kelso, Fletcher, & Lee, 2007). However, even when children fit a generally poor reader profile, their language comprehension abilities may be significantly below their decoding abilities. For example, Bishop and Adams (1990) found that 8-year-olds with SLI obtained significantly lower scores on measures of decoding accuracy and reading comprehension than their typically developing peers. In the SLI group, "reading comprehension scores were disproportionately poor relative to their reading accuracy" (p.

1033) and were better explained by poor language comprehension than by poor reading accuracy. Therefore, while reading fluency difficulties cannot be ruled out, the nature of SLI suggests poor language comprehension as a significant contributor to reading difficulties.

Specific Language Impairment (SLI) and Reading Impairment

The nature of SLI. Children with SLI have persistent difficulties with first language acquisition that manifest in the various domains of language (phonological, semantic, syntactic, and pragmatic). Expressive and receptive language skills may both be affected, or expressive skills may be the principal difficulty (receptive problems alone are rare). The problem is of unknown cause, although genetic factors are believed to be important (Rice, Smith, & Gayán, 2009). To be diagnosed with SLI, children must meet certain exclusionary criteria. They must have normal hearing, no neurological problems, and score within normal limits on nonverbal intelligence tests (Rice, 2009). Thus, children with hearing loss, known neurological conditions (such as epilepsy), intellectual disabilities, or a diagnosis of Autism Spectrum Disorder are excluded from the diagnosis of SLI.

Nonetheless, there is some debate as to whether the problems experienced by children with SLI are specific to language. Some researchers assert that language difficulties are accompanied by cognitive deficits not identified by standard intelligence tests, and others argue that the difficulties are restricted to language (see Silliman & Scott, 2006; Windsor, 2002 for the debate). This debate will not be resolved in the present study. Regardless of the viewpoint regarding the nature of SLI, what is of importance here is that there is consensus that children with SLI exhibit or are at high risk

for problems in reading (Catts, Bridges, Little, & Tomblin, 2008; Catts, Fey, Tomblin, & Zhang, 2002).

Reading difficulties of children with SLI. Several studies have assessed the risk of developing a reading impairment amongst children with SLI. Catts et al. (2002) found that 41.8% of the children in their study who had been diagnosed with SLI in kindergarten were considered to have a reading disability in grade 2 and this proportion remained high at 35.9% in grade 4. These proportions were four to five times greater than the number of students diagnosed with a reading disability in a non-impaired control group. The majority of the students with SLI who did not meet criteria for having a reading disability in second or fourth grades nonetheless scored below average on measures of reading achievement.

Not only are reading difficulties evident early amongst children with SLI, they continue to persist throughout the later school years. Students with SLI who performed significantly below their typically developing peers on measures of reading comprehension in second grade continued to demonstrate a reading comprehension deficit in fourth, eighth, and tenth grades (Catts et al., 2008). In tenth grade, the students with SLI performed at a sixth grade level for reading comprehension, which was significantly below the level of the typically developing group. Clearly, children with SLI are at increased risk for persistent difficulties with reading comprehension. A closer look at the specific skills that contribute to reading comprehension is required in order to gain a better understanding of why these difficulties occur.

Elaborated Two-Path Model of Reading

Although the simple view of reading proposes that reading can be broken down to

just two necessary components (decoding and linguistic comprehension), Hoover and Gough (1990) do not deny that reading is a complex task. Rather, the authors argue that the complexities of reading lie within these two components. That is, both decoding and linguistic comprehension are overarching categories, and each comprises multiple skills that are necessary for successful reading comprehension. Van Kleeck (2008) elaborates the two components of the simple view to provide a more detailed view of the skills required to decode and comprehend text. In this model, inferencing is considered to be a core component of the oral language pathway of reading comprehension; it enables a student to engage in classroom discourse, and facilitates understanding of text. In the literature, inferencing is also referred to as a higher-level cognitive skill; the reader actively engages with the text to derive a meaningful and coherent representation of it (Cain, Oakhill, Barnes, & Bryant, 2001; Trabasso & Magliano, 1996a). More recently, inferencing has been described as being situated at the interface of language and reasoning (Newton, Roberts, & Donlan, 2010). Inferencing is thus a complex skill that draws on both language and cognitive ability.

Model of Text Comprehension

Kintsch's (1988) construction-integration model identifies comprehension of linguistic input, world knowledge, and working memory as crucial to inferencing and to successful comprehension of text. According to this model, the reader draws on a knowledge base composed of networked concepts and propositions that are activated as the text is read. Comprehension occurs in two phases. In the construction process, the first phase of text comprehension, the meaning of concepts and propositions that correspond to the linguistic input (the text) are activated in the reader's knowledge base.

For example, if the word ‘bank’ is read within the text, the concepts that might be activated from within the reader’s knowledge net are ‘bank’ as a financial institution, and ‘bank’ as a riverbank. Associated concepts, such as ‘money’, ‘savings’, ‘mud’, and ‘river’ might also be activated. These concepts may be the ones required to comprehend the text, but might potentially be incorrect given the context of the story. Without having fully processed the sentence or phrase, it is possible that any of the propositions are relevant to the text. Generating multiple propositions from the reader’s knowledge base enables him or her to elaborate on the text; each generated proposition acts as a potential inference. Thus, at the end of the construction process, the reader is left with an enriched but incoherent and potentially contradictory representation of the text.

In the integration phase of the text comprehension process, the information derived in the construction phase is integrated to form a coherent and meaningful text representation. The text is processed in short sentences or phrases at a time. As additional information becomes available, the reader uses his or her working memory to carry information forward so that the next part of the text can be processed and connected meaningfully. In this way, the incorrect or inappropriate propositions and inferences that were activated and generated from the reader’s knowledge base during the construction phase are excluded, and replaced by propositions and inferences that are more appropriate to the context of the text. For example, as the reader processes the sentence, ‘Mary went to the bank to deposit her paycheck’, the concepts, ‘bank’ as a riverbank, ‘mud’, and ‘river’ that were generated during the construction phase would be excluded, and ‘bank’ as a financial institution, ‘money’, and ‘savings’ would be retained. Additionally at this stage, using the propositions that have been derived, specific and

controlled inferences are made by the reader to fill in and causally connect incoherent parts of the text. If the required and correct proposition or elaborative inference was not generated in the initial construction phase, the entire process begins again. The end-result of the integration phase is a coherent and meaningful representation, comprising many relevant inferences that both elaborate on and connect different parts of the text.

Inferencing and Reading Comprehension in Good and Struggling Readers

Inferencing typically develops, as it relates to reading, between the ages of 6 and 12 (Cain & Oakhill, 1998). As was discussed, inferencing has two components. It requires that the reader (1) connect different parts of text using linguistic cues (e.g., by connecting the pronouns 'he' and 'she' with previously-named characters in the text) ("text-to-text" inferences in Cain & Oakhill, 1999; "cohesive" inferences in Bowyer-Crane & Snowling, 2005) and (2) draw upon his or her existing world knowledge in order to fill any gaps in information in the text, or to elaborate on information that is given ("text-to-world" inferences, van Kleeck, 2008). Questions that target a child's inferencing ability can be used to determine how much of the information in a story or text was processed by the child (Wright & Newhoff, 2001). Such questions have shown that children who experience difficulty with reading comprehension have greater difficulty making text-to-world inferences than text-to-text inferences (Bowyer-Crane & Snowling, 2005). Text-to-world inferences will be the focus of the present study.

Text-to-World Inferences

In the literature, text-to-world inferences have been categorized in many ways. In the taxonomy described by van Kleeck (2008), adopted in the present study, there are three text-to-world inference types: causal, informational, and evaluative. Causal

inferences require that the child link two parts of a text while incorporating his or her own world knowledge to fill in any gaps in information. In the context of reading a story, causal inferences link crucial story components (i.e., the initiating event or problem of the story; the protagonist's internal states, goals, and attempts to resolve the problem; the consequences of these attempts; and finally, the solution or outcome of the problem). For an example of causal inferences, a text written for the present study can be used. In the story "Anna and Her Dog" (see Appendix A), Anna is playing fetch with her dog. Given the sentences, "She threw the ball too far. It went into the woods. Anna waited but her dog didn't come back.", a reader might infer, using his or her world knowledge about woods, that Anna was waiting because her dog had gone to fetch the ball, and had not returned because it had gotten lost while searching for the ball in the woods.

The second inferencing category is informational. Informational inferences involve using world knowledge to elaborate information explicitly stated in the text, such as information about setting. Informational inferences create a richer representation of the text, but are not essential to comprehension (Bowyer-Crane & Snowling, 2005; Cain et al., 2001; van Kleeck, 2008). Consider again the story "Anna and Her Dog". In the story, Anna takes her dog on a walk. Based on the story sentence, "They stopped by the swings to play", the reader might use his or her world knowledge about probable locations of swings to infer that Anna is walking her dog in a park.

The third and final type of inference described by van Kleeck (2008) is evaluation. Evaluation enables the reader to use his or her existing values and world knowledge to make judgments of morality, convention, or anomaly about situations in the text. Evaluative inferences are not essential to story comprehension. In the story "Anna

and Her Dog", Anna is looking for her dog who did not come back after going to get the ball in the woods. Encountering the story sentence, "Anna looked until it started to get dark.", the reader might use his or her world knowledge to evaluate Anna's behaviour and surmise that she should not continue to look after dark because she could get lost and might make an evaluative comment to that effect (e.g., "Anna did the right thing to stop looking when it got dark [because she could have gotten lost too]").

Inferencing is typically assessed through the use of comprehension questions following a story. A child's causal inferencing ability is generally assessed by asking 'why' or 'how' questions (Laing & Kamhi, 2002). Using the example given above, the question, "Why didn't Anna's dog come back?" could be asked to assess the reader's causal inferencing. Informational inferences are assessed by asking who, what, when, and where questions (Wright & Newhoff, 2001). For the above example, the reader's informational inferencing could be assessed using the question, "Where did Anna take her dog for a walk?". Evaluative inferences are assessed by asking questions such as "Was that a good thing [for the character] to do?" (van Kleeck, 2008). Given the example from "Anna and Her Dog", the question, "Was it a good thing for Anna to stop looking for her dog when it got dark?" could be asked to assess the reader's evaluative inferencing ability.

Importance of Causal Inferences to Comprehension

Causal inferences have been cited as the most important type of inference for adequate text comprehension (Laing & Kamhi, 2002; Trabasso & Magliano, 1996a; van Kleeck, 2008). Readers who consistently make connections between parts of the text as they read create a coherent and meaningful representation of the text, and are better able

to retain causal information in working memory, thereby increasing the availability of this information when answering questions about the text after reading (Trabasso & Magliano, 1996b).

The majority of inferences made by good comprehenders during reading or listening to a story are causal. Poor comprehenders make significantly fewer causal inferences than good comprehenders while reading (Laing & Kamhi, 2002), as well as on listening tasks, suggesting that the results on the reading task could not be fully explained by poor reading fluency. Cain et al. (2001) similarly found that children with poor reading comprehension responded to comprehension questions requiring causal inferences less accurately than good comprehenders, even when children understood the literal content of stories (as determined through assessment), and world knowledge was controlled for by ensuring all participants had similar knowledge at the outset. This result suggests that content or background knowledge is also insufficient in explaining causal inferencing difficulties in poor comprehenders. Furthermore, Cain and Oakhill (1999) found that causal inferences (along with text-to-text inferences) predicted variance in the reading comprehension of good and poor comprehenders, while measures of literal comprehension did not. Taken together, these findings suggest that the ability to make causal inferences is related to the comprehension difficulties experienced by poor comprehenders and to the superior comprehension performance of good comprehenders (Laing & Kamhi, 2002).

Inferencing in SLI

Both typically developing children and children with SLI have greater difficulty

with inferential than literal comprehension (Wright & Newhoff, 2001[1]). For children with SLI, this pattern has been shown on listening comprehension tasks (Botting & Adams, 2005; Norbury & Bishop, 2002; Wright & Newhoff, 2001), and even when stories contained familiar content (Norbury & Bishop, 2002), and vocabulary was at a level that all children could comprehend (Wright & Newhoff, 2001). These results are consistent with those reviewed above for poor comprehenders, and suggest that the text comprehension difficulties experienced by children with SLI do not arise only because of problems with decoding, reading fluency, or inadequate world knowledge.

Children with SLI also perform below their typically developing, chronological age-matched peers on both literal and inferential comprehension, even when the reading material is below grade level (Norbury & Bishop, 2002). Performance on comprehension questions, asked after reading a text or listening to stories, is comparable to younger children who have been matched on some aspect of expressive and/or receptive language ability (Adams, Clarke, & Haynes, 2009; Bishop & Adams, 1992; Botting & Adams, 2005; Norbury & Bishop, 2002; Wright & Newhoff, 2001). The resemblance of children with SLI to younger, language-matched children on comprehension tasks can be interpreted in two ways: (1) the difficulties with comprehension arise because of difficulties with oral language, as the groups are matched on language ability, or (2) the comprehension difficulties reflect delays in cognitive development in the SLI group (Newton et al., 2010; Wright & Newhoff, 2001). Given the debate as to how specific SLI actually is to language (e.g., Silliman & Scott, 2006) and the uncertainty as to whether

[1] In the literature, as well as in practice, impairments that appear specific to language are referred to with various labels, including language-learning disability, language disorder, and language impairment. For the sake of clarity and consistency, the term SLI has been used in this report to identify such children.

inferencing is best considered an oral language skill, a cognitive skill, or a combination of both (Cain et al., 2001; Newton et al., 2010; van Kleeck, 2008), it is difficult to disentangle these explanations.

However, Newton et al. (2010) support a linguistic explanation for the inferencing difficulties observed in SLI. In their study, children in an SLI group scored significantly lower on an inferencing task presented pictorially and verbally than a group of typically developing, age-matched controls, but performed similarly to a language-matched group. Scores on a receptive grammar measure contributed to the between-group differences, while measures of both working memory and cognitive ability did not. Given that these results were obtained when the task was presented both orally and pictorially, the authors suggest that language mediates performance even on non-verbal inferencing tasks.

It has been documented that children with SLI have difficulty with text-to-world inferential comprehension. However, the performance of children with an SLI profile (described by the authors as 'language-learning disabled') on comprehension questions in terms of inference type has been assessed in only one study (Wright & Newhoff, 2001). Consistent with the group differences observed for overall inference scores (alluded to in the preceding paragraph), the children with SLI performed significantly below the chronological age-matched group and at a similar level to the language-matched group on questions requiring informational and causal inferences (the authors refer to literal statements as "premises", informational inferences as "value inferences", and causal inferences as "logical inferences"). All three groups performed significantly better on informational questions than on causal questions (Wright & Newhoff, 2001).

The finding that causal inferences are the most difficult to make for all children

suggests that inference types emerge at various points in development (Cain & Oakhill, 1999). Making a causal inference requires that the child (1) connect two premises in the text, and (2) incorporate his or her own world knowledge to create meaning. The other forms of text-to-world inferences require that the child draw on his or her world knowledge to elaborate on or make sense of only one premise in the text. Given the greater demands of making a causal inference, it is plausible that a child is only able to successfully make this type of inference later in development.

In the studies reviewed thus far, both inferential and literal comprehension have typically been assessed by asking children comprehension questions *after* they read or listen to a story. Another procedure has been used to reveal the inferences children make *during* reading: the think-aloud procedure.

The Think-Aloud Procedure

In a research context, a 'think-aloud' refers to individuals' verbal reports of their thoughts as they enter and are processed in working memory, either during or immediately following a task (Crutcher, 1994). Think-alouds have been used to investigate a wide range of cognitive processes, including the number and type of inferences children make during reading (Laing & Kamhi, 2002; Trabasso & Magliano, 1996b). Children read one sentence of a passage at a time and are instructed to verbalize whatever thoughts come to mind after reading each sentence. Given that the child is not instructed to think about the process of reading or what his or her thoughts might mean, the verbalizations are thought to expose the information that the child is currently processing in working memory, and that is consciously available to him or her while reading (Ericsson & Simon, 1993).

The think-aloud procedure has revealed marked differences in the overall number of verbalizations made by good and poor adult readers (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Moreover, adults who made many verbalizations during the procedure showed better comprehension for what they had read than adults who made few (Chi et al., 1989). Trabasso and Magliano (1996b) extended the think-aloud procedure to children in third grade to investigate inferencing. They found that 70% of verbalizations made by third graders while reading were inferences, but that the children made fewer causal inferences than adults. Laing and Kamhi (2002) partially replicated the study and found that children characterized as poor comprehenders made significantly fewer causal inferences than good comprehenders during a think-aloud.

Successful reading occurs when a reader has processed a text and has made specific and deliberate causal connections in order to make sense of the text (Kintsch, 1988). It is thought that readers who aim to make causal connections between parts of text by linking propositions are more likely to hold meaningful information in working memory, thereby enabling them to retrieve and use the information after reading to correctly answer comprehension questions or to recall text (Trabasso & Magliano, 1996b). That is, the causal inferences that good comprehenders make during reading contribute to their superior performance on comprehension questions and recall tasks, while the difficulty poor comprehenders appear to have in making this type of inference might similarly affect their performance on such tasks. Information about the number and types of inferences a reader makes is critical in understanding a child's reading difficulties. This is information that can not be obtained through the use of off-line comprehension questions.

Although there are clear benefits to using a think-aloud procedure, there are three important limitations to be considered. First, a child might think more about the text than he or she might normally due to task demands and prompting. Second, there are increased demands placed on the child as he or she must simultaneously perform the task and report on it, thus breaking up the task of reading, however briefly. Third, children with SLI may have difficulty expressing their thoughts. Despite these limitations, it has been argued that the think-aloud procedure provides valuable and rich insight into a child's reasoning process that can not be obtained through the use of comprehension questions alone (Phillips, 1988). Additionally, previous research assessing the storytelling ability of adolescents with SLI (Wetherell, Botting, & Conti-Ramsden, 2007), and the comprehension of expository text of children with SLI, using a think-aloud procedure (Gillam, Fargo, & Robertson, 2009), suggests that children in upper elementary school (as in the present study) should be able to meaningfully express their ideas during a think-aloud procedure using narrative texts. Moreover, the successful use of the protocol with poor comprehenders, all of whom were shown to have previously unidentified difficulties with oral language, (Laing & Kamhi, 2002), suggests that the results will be informative.

The Present Study

The goals of the present study were to (a) compare the inferencing abilities of children with SLI and with typical language development (TLD), as assessed by offline comprehension questions, (b) to extend the research by investigating online inferencing of children with SLI using a think-aloud procedure, and (c) to determine whether a think-aloud procedure facilitates reading comprehension of narrative texts for children with SLI

and for age-matched peers with TLD.

A group of children with SLI and an age-matched group of children with typical language development (TLD) were compared on their responses to comprehension questions following stories read silently by each child and simultaneously heard over a speaker during an individual session (the read-through condition). It was hypothesized that (1) children with SLI would perform below the TLD group on both literal and inferential questions (2) both groups would have greater difficulty with inferential than literal comprehension questions, and (3) both groups would have lower scores on questions requiring causal inferences than on questions requiring informational inferences.

The second objective was to compare children with SLI to the TLD group in terms of the inferences they verbalized after simultaneously reading and listening to each sentence of a story (the think-aloud condition). Based on the findings by Laing and Kamhi (2002) in a study conducted with good and poor comprehenders, the hypotheses were (1) the SLI group would make fewer and less accurate inferences than the TLD group and (2) the SLI group would make fewer causal inferences than the TLD group.

A third objective was to compare participants' performance on inferential and literal comprehension questions across the two conditions (think-aloud and read-through). It was hypothesized that performance would be enhanced in the think-aloud condition for both the SLI and TLD groups. Although the premise of the think-aloud procedure is that it will reveal what a participant is thinking, Laing and Kamhi's (2002) results suggest that children might make more inferences in the think-aloud condition than they would spontaneously, thereby facilitating reading comprehension.

Finally, measures of oral language served to distinguish the SLI and TLD groups. The oral language scores, as well as scores on tests of nonverbal cognition, were additionally examined as correlates of children's performance on the reading comprehension questions and the total number of causal inferences made during reading (that is, in the think-aloud condition).

Method

Research Design

The design is a 2 x 2 mixed model, in which the between-group factor was group (SLI or TLD), and the within-group factor was reading condition (read-through or think-aloud). Each participant in the SLI and TLD groups participated in the two reading conditions: read-through and think-aloud.

Participants

Pilot participants. As reported in the procedures below, the stories, comprehension questions, and the think-aloud procedure were piloted with adults and children in three phases. Eight adults participated in phase 1 (*M* age in years and months = 29;8, *SD* = 2;1) and 12 adults (*M* age = 26;10, *SD* = 4;8) participated in phase 2. All adults had a minimum of 2 years of university-level education. Two children (one girl and one boy), aged 7;10 and 9;11 respectively, participated in phase 3 piloting. All pilot participants were recruited from the community by word-of-mouth.

Study participants. Although ethical approval was obtained from the university and a local school board to recruit in the community and in local schools, this came at the end of the school year and entry to schools was not possible. Therefore, participants were initially recruited within the Greater Montreal area during the summer months

through flyers posted in local settings frequented by children and families (e.g., public libraries), word of mouth, and letters and flyer distribution to local speech-language pathologists. Four participants, all within the TLD group, were recruited from the community; attempts to recruit children with language impairment were unsuccessful. In the fall of 2011, 23 participants were recruited from local elementary schools with the assistance of classroom teachers, who sent information letters and consent forms home to parents in the students' agendas (the usual form of communication between teachers and parents), in keeping with the processes recommended by the university and school board ethics committees.

Participants recruited for the SLI group were 13 upper elementary students who had been identified as having a mixed (expressive and receptive) language impairment by speech-language pathologists and were regularly receiving speech-language services. For 11 of the 12 participants, the impairment was severe, according to criteria set by the provincial Ministry of Education [Ministère de l'Éducation, du Loisir et du Sport] (Gouvernement du Québec, 2007). Eight boys and 5 girls were originally recruited. One female participant was dropped due to factors compromising the validity of language test scores (i.e., English as a second language with no English spoken at home). The final sample was comprised of 12 participants (*M* age = 10;5, *SD* = 0;8), including 8 boys (*M* age = 10;6, *SD* = 0;7) and 4 girls (*M* age = 10;3, *SD* = 0;11). A greater proportion of boys than girls participated in the study. This was anticipated given the distribution of boys and girls with language impairment found on clinical caseloads, as well as research findings of a greater prevalence of SLI among boys (Rice, 2009). The participants in the 'typical language' group were, by design, matched by sex and age. All participants were

receiving educational instruction in English, and reported speaking English the majority of the time at home. All, however, had some exposure to a second or a third language, either at school through second language instruction, or in their community or home.

Sixteen students with typical language development, identified by teachers and parents as having average or above average language and reading ability were recruited for the TLD group. Further inclusion criteria were that the participant had not repeated a year in school, and received English instruction in school at least 50% of the time (it is not uncommon for English-speaking children in the province to be enrolled in French immersion education at least half-time). As in the SLI group, all participants reported speaking English the majority of the time at home with some exposure to a second or third language. Testing was discontinued after the first session for two participants: one boy who had previously repeated a year of school, and one girl who had only recently immigrated to Canada. A further two participants recruited for the TLD group were dropped from the sample: one boy whose standard scores were one standard deviation or more below the mean on all standardized measures, and one girl who performed below one standard deviation on the language test. The final sample size was 12 (*M* age = 10;3, *SD* = 0;6) with 8 boys (*M* age = 10;4, *SD* = 0;6) and 4 girls (*M* age = 10;2, *SD* = 0;7) in the TLD group.

Although the Test of Nonverbal Intelligence-3 (TONI-3) was not used as a screening measure for group inclusion, it was used to rule out intellectual impairment for both groups. The mean scores for both groups were within average range (within one standard deviation of the mean) and not significantly different between groups. It should be noted, however, that four participants in the SLI group had a standard score below 85,

as did five participants in the TLD group. In the SLI group, two of the four participants who scored below 85 scored within one standard error of measurement (80 to 84), one of the participants scored within two standard errors of measurement (75 to 79), and one participant scored within three standard errors of measurement (70 to 74). These participants were kept in the sample, given that each child had a diagnosis of language impairment prior to the present study that excluded intellectual impairment based on more extensive testing. Five participants in the TLD group had a score between 80 and 84, within one standard error of measurement of a standard score of 85. Lower scores may have, in some cases, reflected test anxiety, particularly for children in the TLD group who may have had less experience with formal testing, and especially as the TONI-3 was the first measure to be administered in the first session (see procedures regarding task order).

Measures and Apparatus

The following measures and apparatus were used to assess students' oral language, nonverbal cognitive ability, nonverbal working memory, and reading comprehension and inferencing ability.

Apparatus. All tests requiring verbal responses from the participants were audio recorded using a Panasonic RR-US-570 digital recorder. All story stimuli were pre-recorded into mp3 files and saved onto an iPod Touch mp3 player. A portable X-Mini II capsule speaker was used during testing to play the story recordings to the participants.

Clinical Evaluation of Language Fundamentals-Fourth Edition. The CELF-4 (Semel, Wiig, & Secord, 2003) is a standardized, norm-referenced test designed to assess the language abilities of individuals aged 5 to 21 years. The CELF-4 has excellent

reliability and validity and is also one of the most frequently used measures used by clinicians to diagnose SLI (Eickhoff, Betz, & Ristow, 2010). The standardization sample consisted of 2,650 students, and was representative of the 2000 U.S. population for age, gender, race/ethnicity, socio-economic status (based on education level of the primary parent), geographic location, and prevalence of developmental disabilities (language-based or other). For composite scores, internal consistency (measured using Cronbach's alpha) was between .87 and .95 and split-half reliability ranged from .87 to .95. Inter-scorer agreement for subtests ranged from .88 to .99. The CELF-4 also has high sensitivity and specificity, which is important when making diagnoses of language impairments and, for the purpose of the present study, to identify children as having typical language development or a language impairment.

The Receptive and Expressive Language subscales were administered to assess each student's oral language ability. The Receptive Language subscale includes two tasks: Concepts and Following Directions, where the child is asked to point to pictures that correspond to instructions varying in length and complexity, and including different concepts (for example, spatial and temporal concepts) and Word Classes-Receptive, where the child listens to four words and is asked to choose two semantically-related words. The Expressive Language subscale is made up of three tasks: Recalling Sentences, in which the child is asked to repeat verbatim a sentence that is read by the examiner; Formulated Sentences, in which the child is told a word and is asked to make up a sentence that is contextually relevant to a picture; and Word Classes-Expressive, in which the child must explain how two words are related in meaning. The Receptive and Expressive scales yield standard scores that are combined in different ways to form Core

Language and Language Memory composites (referred to as indices by the test authors).

Test of Nonverbal Intelligence-3. The TONI-3 (Brown, Sherbenou, & Johnsen, 1997) is a standardized, norm-referenced measure of nonverbal reasoning and problem-solving ability that can be used with individuals between the ages of 5 and 86 years. The TONI-3 was normed on a sample of 3,451 children and adults in the U.S. in 1995 and 1996. The sample was representative of the U.S. population in terms of geographic region, gender, race, urban/rural residence, ethnicity, socio-economic status (based on family income and level of educational attainment of adult participants and parents of minor participants), and identified disability.

The TONI-3 consists of two highly correlated, parallel forms (forms A and B). Alternate forms reliability for children aged 9 to 11 years is between .79 and .85. Interscorer reliability for the TONI-3 was .99. Test-retest reliability, completed for participants aged 13 years and older, was greater than .89 for both forms. According to the test authors (Brown et al., 1997), TONI-3 scores correlate with the overall nonverbal IQ on the CTONI (Hammill, Pearson, & Wiederholt, 1996; as cited in Brown et al., 1997), r = .76 and .74 for Forms A and B, respectively; the full scale IQ of the WISC-3 (Wechsler, 1991; as cited in Brown et al., 1997) r = .63 for both forms; and the full scale IQ on the WAIS-R (Wechsler, 1981; as cited in Brown et al., 1997), r = .73 and .71 for Forms A and B, respectively.

Data provided by the authors as evidence of construct validity includes differences in nonverbal cognitive ability at different ages (correlation coefficients of .63 for Form A and .60 for Form B, between 6 and 18 years of age), correlation of scores with school achievement (coefficients of .55 to .76 for students with learning disabilities

between the ages of 8 and 16 years, as well as an average correlation of .59 with several measures of achievement), and differences in scores between groups known to display differences in nonverbal cognitive ability (Brown et al., 1997).

Each form contains 50 items, arranged in order of difficulty. Each item consists of a matrix of symbols in which one symbol is missing (See Appendix B for sample item). For each item, the displayed and missing symbols are constrained by a specific relationship (e.g., matching, analogous, additive, progressive). The examiner administers the test demonstration items nonverbally (with a pointing and questioning gesture), and the participant responds by pointing to one of 4 or 6 options that would best complete the matrix. A ceiling is reached, and testing is ended, when the participant makes three errors out of five consecutive items. As oral language is not required to conduct the test, the TONI is recommended for use with language-impaired individuals (McGhee & Lieberman, 1990).

Working Memory Test Battery for Children. The Working Memory Test Battery for Children (WMTB-C) (Pickering & Gathercole, 2001) is a standardized, norm-referenced measure for children aged 5 to 15 years that comprises nine subtests of verbal and nonverbal working memory. The WMTB-C was standardized on a sample of 729 British children (included 98 students with special educational needs) aged 4;7 to 15;9 years old, selected from schools in both rural and urban settings.

Two nonverbal tests of working memory, the Block Recall and Mazes Memory subtests, were selected for use in the present study. These subtests are highly correlated with WMTB-C verbal working memory tasks of Counting Recall (r = .60 and .65, respectively) and Listening Recall (r = .55, .59, respectively). However, the nonverbal

tasks avoid the confound of oral language believed to negatively affect the performance of children with SLI (Montgomery, Magimairai, & Finney, 2010), and children with reading difficulties (Savage, Lavers, & Pillay, 2007) on verbal working memory tasks. Test-retest reliability for the Block Recall and Mazes Memory tasks, for children aged 9;7 to 11;4 years, was .43 and .53, respectively. The Block Recall and Mazes Memory subtests were significantly but not highly correlated with one another (r = .30). The test authors considered inter-rater reliability unnecessary for either the Block Recall or the Mazes Memory tasks, as there was no subjective decision-making necessary with regards to an item being correct or incorrect on either subtest.

In the Block Recall task, the experimenter taps a series of blocks, after which the child is required to repeat the sequence. In the Mazes Memory task, the participant watches as the examiner traces a path through a maze with her finger. The participant is then asked to draw the same path on an identical maze in a test booklet. Both tasks increase in complexity as the participant correctly replicates the demonstrated items, and testing is stopped when the participant makes three errors out of a block of six trials. Standard scores for both the Block Recall and Mazes Memory subtests were obtained.

Experimental stories and reading comprehension questions.

Picture stories. Two picture stories, adapted from McKeough et al. (1996), served to measure children's volubility on an oral narrative task - a task that was considered to be similar, yet less demanding, than the think-aloud task. The storytelling also served as a warm-up activity before beginning the reading activities. Each story consisted of a sequence of five simple, black and white line drawings. One story depicted a young boy falling off of his bike and calling for help, and the other showed a

young girl hitting a home run during a baseball game (see Appendix C for picture stories).

Think-aloud demonstration stories. One demonstration story (Rachel and the Giant, 51 words) was adapted from Trabasso and Magliano (1996a) with permission (J. Magliano, personal communication, January 12, 2011). A second demonstration story (Ivan and the Dragon, 55 words) was written by the researcher to parallel the first story (see Appendix D). Stories were written at a second and third grade level respectively, as measured by the Spache Readability Formula, which takes into account the difficulty of words used and average number of words per sentence (Spache, 1953). Participants read the story from a printed copy in front of him or her and simultaneously listened to audio recordings of the stories, told at a normal speaking rate and played over a speaker at a conversational volume. Demonstration story themes were fantastical in nature (a young girl scaring a giant away from her town, and a young boy befriending a dragon), which contrasted with the realistic nature of the experimental stories so as to minimize carryover effects from the demonstrations to the children's own think-alouds.

Experimental stories. Four stories, one from Bishop and Adams (1992) adapted with permission (D. Bishop, personal communication, December 23, 2010), and three written by the experimenter (see Appendix A), were presented to each participant to measure story comprehension and inferencing. Stories were audio recorded at a normal speaking rate and regular conversational volume to maintain consistency during testing. The stories ranged in length from 108 to 114 words long and were intentionally written at a readability level below the participants' grade levels to avoid the potential effects of vocabulary knowledge, syntactic complexity, or decoding difficulties on the variables of

primary interest. According to the Spache Readability Formula, the stories were at a second grade level (Spache, 1953).

Story themes were selected to ensure that the story content would likely be familiar to all students. Two of the stories centered on the protagonist's goal of making something (a go-cart and a birthday cake), while the other two stories centered on the protagonist experiencing a difficult event (falling through the ice while skating and attempting to find a lost dog).

Comprehension questions. A set of nine comprehension questions was written for each story. Each set included three literal, three causal inferencing, and three informational inferencing questions. Literal questions were based on information that was stated explicitly within the text. Inferential questions were broken down into two categories: informational and causal. Informational questions required that the participant use his or her world knowledge to expand on a single, explicitly stated proposition, while causal questions required that the participant use his or her world knowledge to identify information that was implicitly stated in the text in order to connect two explicitly stated propositions.

Procedure

Stage 1 piloting. The experimental stories and comprehension questions were piloted in the read-through condition with eight adult participants. Piloting took place individually, in either a quiet room at Concordia University or in the participant's home. The researcher read all four stories out loud while the participants silently read a printed version of the story. Each story was immediately followed by a set of comprehension questions about the story. Participants were given the option of taking a break halfway

through the session. Before reading each story, the following instructions were given: "Now we are going to read a story. I am going to read it out loud, and you will read along in your head. Then, I am going to ask you some questions about the story, so do your best to understand it while we're reading." After reading the story, the participants were told, "Now I am going to ask you some questions about the story. It's okay to guess at the answers." The printed story was moved out of the participant's view during the question period. If no response was given, or participants indicated that they either could not recall or that the requested information was not presented in the text, the prompt, "Do you want to guess?" was used. Participant feedback about inconsistencies or areas of difficulty in the stories and comprehension questions was used to make adjustments to the stimuli, and responses to the comprehension questions were used to develop a comprehension question scoring key.

Stage 2 piloting. Once all adjustments had been made to the stories and comprehension questions, the read-through and think-aloud conditions were piloted with 12 adult participants. For the read-through, the Stage 1 piloting procedures were replicated. Additionally, stories were counterbalanced across condition, but order of condition alternated and was kept constant (for each participant, each session began with a story in the read-through condition, followed by a story in the think-aloud condition, and so on).

Before reading the first think-aloud story, the researcher provided the participant with a scripted demonstration using the "Ivan and the Giant" story (later modified to "Rachel and the Giant" to avoid gender stereotypes; see Appendix E for demonstration story scripts). As all stories were read in one session, a second demonstration was not

given before the second story in the think-aloud condition. Before both think-aloud stories, participants were instructed, "Now we are going to read another story. This time we are going to do something a little different. We are going to read the story one sentence at a time. I will read the story out loud, and you will follow along in your head. After each sentence, I will give you a thumbs-up, and you will explain what you understand about the story." Before the first think-aloud story, participants were further instructed, "First, I'll show you how to do it, and then it will be your turn." Following the demonstration, the instructions were repeated. As in the read-through condition, participants were told that they would be asked questions about the story once they had read it, and to do their best to understand the story.

In both conditions, stories were presented to participants printed one sentence per line, double-spaced, on an 8.5 x 11" piece of white paper. In the think-aloud condition, only one sentence of the story was visible to the participants at a time while they were explaining their understanding of the story. The rest of the story was covered up using a piece of white paper. Participant responses during the think-aloud procedure were audio-recorded for later transcription. As in the read-through condition, the story was placed out of the participants' sight during the question period. Participants were once again told that they would now be asked some questions about the story, and that it was okay to guess at the answers.

Participant responses to comprehension questions were used to further modify the comprehension question scoring key, and statements made in the think-aloud conditions were used to develop a think-aloud scoring system.

Stage 3 piloting. Following stage 2 piloting, a second demonstration story was

written by the researcher (Ivan and the Dragon; see Appendix D). The final demonstration and experimental stories were printed, one sentence per page, in a story stimulus book (see Appendix F for example). All stories were audio-recorded by the researcher into mp3 files, and played over a speaker during testing. These stimuli were piloted in both the read-through and think-aloud conditions with two children. As in the stage 2 piloting, participants read all four stories (two stories in each condition) and answered corresponding comprehension questions in a single session. A break was taken after reading the first two stories, in which the researcher engaged in a short origami airplane activity with the participants. The testing procedure in this stage of piloting was identical to the Stage 2 Piloting procedure. The second think-aloud demonstration story, Ivan and the Dragon, was piloted with one of the participants, whereas the Rachel and the Giant demonstration story (see Appendix E for demonstration story scripts), was piloted with the other participant.

Study testing. Individual testing was conducted in a quiet area outside of the classroom or in the participant's home. Test sessions were, on average, three days apart (range 1-7 days) and each lasted approximately one hour. Participants were tested in both sessions by one of two examiners, either the researcher or a trained research assistant. Instructions for experimental tasks were scripted to maintain consistency across examiners (see Appendix G for test prompts), and standardized procedures were followed for the norm-referenced measures. Tasks requiring verbal responses were audio-recorded for later review. During administration of nonverbal measures, responses were scored by hand during the session on scoresheets provided with the tests, and the sheets were out of sight of the participant.

Verbal and nonverbal measures were distributed across the two sessions to minimize fatigue or boredom of participants. Order of test administration was kept constant for all participants within test sessions. Administration of the story measures was identical to the testing procedures in Stage 3 Piloting, with one story in each condition, for a total of two stories read per session. The four stories were counterbalanced across the think-aloud and read-through conditions for both groups, to ensure that each story was read the same number of times in each condition for both groups.

At the start of each session, the examiner read an assent script (see Appendix G), and verbal assent was obtained from the participant. Session 1 began with the TONI-3, followed by the biking picture story narrative, a story in the read-through condition, the Rachel and the Giant think-aloud demonstration, and a story in the think-aloud condition. After administering the reading measures, the participants were given a short break in which they made colourful paper airplanes from an origami airplane kit, followed by administration of the Word Classes-2 and Recalling Sentences subtests of the CELF-4. Session 2 began with the Block Recall, and then the Mazes Memory subtests of the WMTB-C. Administration of the reading measures began with the baseball picture story narrative, followed by the Ivan and the Dragon demonstration story, a story in the think-aloud condition, and a story in the read-through condition. Following administration of the story measures, the participant was given a short break in which they made a second origami airplane. After the break, the Concepts and Following Directions and Formulated Sentences subtests of the CELF-4 were administered. Participants were thanked for their participation in the study and given the paper airplanes they had made

during their break. Gifts were not thought to be appropriate in the school setting, as not all students in the classrooms from which participants were recruited were selected to participate, as some had not returned signed consent forms, and others did not meet inclusion criteria for the study.

Scoring

Standardized measures. Standard scores were obtained for the TONI-3, the Core Language[2], Receptive Language, Expressive Language, and Language Memory composites of the CELF-4, and the Block Recall and Mazes Memory subtests of the WMTB-C.

Picture story scoring. Student's picture story narratives were scored by a research assistant. The total number of independent clauses provided in each narrative were counted from the audio recording (ignoring reformulations and utterances that trailed off).

Comprehension question scoring. Students' responses to comprehension questions were scored according to a 3-point scoring system adapted from Bishop and Adams (1992). Two points were awarded for a correct response, 1 point was awarded for a response that was correct but that omitted an important detail, and 0 points were awarded for no response or responses that were incorrect (See Appendix H for Experimental story scoring keys). The maximum total score that could be obtained was 18 (9 questions per story x 2 points maximum) and the minimum was 0. Two

[2] One participant in the current study was 8 years, 11 months. For children of this age, the Core Language composite excludes one of the subtests (Word Classes, on which standard scores can nonetheless be calculated for individuals between 5 and 8 years) that form the composite for individuals aged 9 years or more, replacing it with Word Structure. Given that the Word Classes and Word Structure subtests tests are not equivalent, and for the sake of consistency with other participants, the Core Language composite for this participant was calculated as it would be for children aged 9 years and over.

adjustments were made after reviewing the comprehension questions during scoring: one literal question in the story "Andrew goes Skating" was re-categorized as informational, and one informational question in "Gabby and the Stroller" was removed, as the wording may have led to some confusion in the children's responses. These adjustments did not change the analyses, in that the planned proportional scores (total points in a category/total possible points in that category * 100) could still be calculated.

The researcher scored 100% (n = 96) of all comprehension questions. A research assistant, blind to group and purpose of the study, scored 33% (n = 32) of the questions. The 32 questions were selected randomly from a set that equally represented the groups and were drawn equally from the four stories. Interscorer reliability was 94.61% (weighted *kappa* = .91).

Think-aloud transcription and scoring. To analyze the inferences the children made during the think-aloud procedure, each child's statements during the think-aloud condition were transcribed into Microsoft Word files by the researcher (see Appendix I for a sample transcription). Think-aloud statements were segmented into clauses that expressed a main idea, which was generally indicated by the presence of a main verb (see Appendix J for segmenting rules). Segmenting rules were devised to ensure that statements encompassed a main idea while allowing for natural variations in spoken language, such as ellipses and omissions. Once transcribed, think-aloud statements were categorized as literal or inferential. Literal statements consisted of full and partial repetitions and paraphrases of the current story sentence. Inferential statements were further categorized as causal, informational, or evaluative. If semantic, phonological, and/or syntactic errors made it impossible to interpret the statement as either literal or

inferential, it was coded as uninterpretable.

Given that a single story sentence could generate an infinite number or a 'chain' of statements, either relating directly to the current story sentence, to a previously read story sentence, or to the participant's preceding think-aloud statements, a coding sequence was integrated to the scoring procedure (see Appendix K) to ensure that codes were systematically assigned to statements. Using the scoring procedure, the coder categorized each think-aloud statement as literal, causal, informational, evaluative, or uninterpretable. If the statement was either a full or partial literal repetition or paraphrase of the current story sentence, it was coded as literal. If the statement was not a literal statement or if it contained added information, the coder moved on to consider whether the statement communicated cause or effect in relation to the current story sentence, a preceding story sentence, or a previous statement. If the statement did not fit the causal category, the coder moved on to the informational category. If the statement answered 'what, how, who, where, when' questions about either the current story sentence, a preceding story sentence, or a previous statement, it was coded as informational. If the think-aloud statement could not be coded as informational, the coder then considered the final inferential category, evaluation. If the statement was used by the reader to judge or evaluate a state, event, action, or object described by the focal sentence or a previous statement based on the reader's world knowledge, social conventions, or personal values, it was coded as evaluative. Finally, if the coder could not interpret the statement due to semantic, phonological, and/or syntactic errors, the statement was coded as 'uninterpretable'.

In addition to being categorized, the children's statements were scored using a 3-

point scoring system (0, 1, or 2 points). Literal statements were scored 2 points if they were an exact repetition of the focal sentence or a close paraphrase; 1 point if they were partially correct but included grammatical or semantic errors that led to a significant change in meaning or rendered part of the statement ambiguous; and 0 points if the statements were incorrect or did not make sense. A 3-point system was also used to score inferential statements. An inferential statement was scored 2 points if it was relevant and logically linked to the text or to a previous statement, and made sense given the context of the story; 1 point if the statement was relevant but an important piece of information was omitted that led to a change in meaning, or if the statement was inaccurate but still related to the story; 0 points if the statement seemed outlandish or nonsensical given the information provided in the story. The minimum total score possible was 0 but there was no maximum, as an infinite number of statements could be generated by participants and scored. Statements coded as uninterpretable were not given a score, and were excluded from analyses.

The total number of literal, causal, and informational statements produced during the think-aloud procedure was calculated. Additionally, the number of causal statements that were scored 1 point and 0 points were calculated as a proportion of the total number of statements in that category (i.e., (number of causal statements scored 1 point / total number of causal statements) * 100) to allow for between-group comparisons of statement accuracy.

The researcher scored 100% (n = 48) of the think-aloud transcriptions. To calculate interscorer reliability, a trained research assistant who was blind to group scored 25% (n = 12) of the transcriptions. The transcriptions were selected randomly from a set

that equally represented the groups and the four stories. Interscorer reliability of 85.44% (*kappa* = .72) was obtained for statement coding. Reliability was established for only the literal, causal, and informational categories; there were too few instances of the evaluative codes overall to gather meaningful reliability data and moreover, the category was not further analyzed here.

For the assignment of points to the statements, reliability was 92.90% (weighted *kappa* = .31, *SE* = .09) (weighted kappa is appropriate for ordinal scores). Given the large discrepancy between the % agreement and the kappa value, disagreements between coders were reviewed by the research supervisor and researcher. Most disagreements reflected a tendency for the second coder - the trained assistant - to attribute 2 points to 1 point responses, an issue returned to in the discussion.

Results

Normality tests. The Shapiro-Wilks test was used to examine normality for all variables, and the results were additionally considered in light of visual inspection of scatterplots and histograms. When these tests indicated violation of the normality assumption, scores were transformed using transformations (consistently, square root). When violations of normality remained despite the transformations, nonparametric statistical tests were used: the Mann-Whitney U to assess between-group differences and the Wilcoxon Signed Ranks to assess within-group differences (there is no non-parametric equivalent to a repeated measures ANOVA with a between-group factor, though such an analysis would have been ideal given the study design).

Age. The children ranged in age from 8;11 to 11;3. The mean age of the SLI group was 10;5 years (*SD* = 0;8) and the mean age of the TLD group was 10;3 years (*SD*

= 0;6). An independent *t*-test showed that the mean group ages were not significantly different, $t(22) = -.562$, $p = .580$.

Language measure (CELF-4). Scores on the subtests of the CELF-4 (described in the Method) can be combined to yield various composite standard scores: a Receptive Language score, an Expressive Language score, a Core Language Score comprised of the Receptive and Expressive scores, and a Language Memory score entailing all but one of the subtests involved in the Core Language score. These scores are reported in Table 1 for descriptive purposes, along with standard scores for all measures. Given that the Core Language and Receptive Language scores were not normally distributed, they were transformed using a square root transformation. Independent *t*-tests were conducted on the transformed scores for these variables. The SLI group performed significantly below the TLD group on each of these composites: Receptive Language (SLI $M = 8.38$, $SD = .35$; TLD $M = 9.99$, $SD = .87$), $t(22) = 5.965$, $p < .001$; Expressive Language (SLI $M = 63.83$, $SD = 6.85$; TLD $M = 110.83$, $SD = 14.84$), $t(22) = 9.960$, $p < .001$; Core Language (SLI $M = 7.79$, $SD = .48$; TLD $M = 10.23$, $SD = .70$), $t(22) = 10.005$, $p < .001$; and Language Memory (SLI $M = 57.17$, $SD = 6.67$; TLD $M = 101.92$, $SD = 14.48$), $t(22) = 9.726$, $p < .001$.

Non-verbal measures (WMTB-C and TONI-3). For the Mazes Memory subtest of the WMTB-C, the performance of the two groups was not significantly different (SLI $M = 94.33$, $SD = 16.73$), $t(22) = .51$, $p = .62$. However, the SLI group ($M = 81.25$, $SD = 16.73$) performed significantly below the TLD group ($M = 96.67$, $SD = 16.73$) on the Block Recall subtest, $t(22) = 2.26$, $p = .035$.

For the TONI-3, a Mann-Whitney U test revealed no significant difference

between groups, $U = 61.50$, $z = -.61$, $p = .54$. The untransformed standard scores are displayed in Table 1.

Table 1

Descriptive Statistics for Language and Nonverbal Measures, by Group

	SLI Group ($n = 12$)		TLD Group ($n = 12$)	
	M	*SD*	*M*	*SD*
Measure				
CELF-4: Receptive Language	70.25	5.79	100.58	18.12
CELF-4: Expressive Language	63.83	6.85	110.83	14.84
CELF-4: Core Language	60.92	7.40	107.42	19.37
CELF-4: Language Memory	57.17	6.67	101.92	14.48
WMTB-C: Block Recall	81.25	16.73	96.67	16.73
WMTB-C: Mazes Memory	94.33	18.79	98.42	20.76
TONI-3	89.67	11.99	96.08	17.90

Note. Standard scores reported with a population mean of 100, $SD = 15$.

Picture stories. The number of independent clauses made by the TLD group in the picture story narratives was not normally distributed. Mann-Whitney U tests were conducted to analyze the number of independent clauses produced during picture story narratives for both stories individually. The alpha level was adjusted for multiple comparisons: Bonferroni adjusted alpha = .025 (.05/2 comparisons). Results were not significant for the first narrative, $U = 40.50$, $z = -1.89$, $p = .029$ (one-tailed), nor for the second narrative, $U = 72.00$, $z = .000$, $p = .500$ (one-tailed).

Comprehension questions. Four comprehension question variables were not normally distributed (read-through literal, read-through informational, think-aloud literal,

and the total scores in both conditions). While the remaining comprehension question variables were normally distributed, non-parametric tests were used to allow for comparison of results across question type: Mann-Whitney U tests for between-group comparisons, and Wilcoxon Signed Ranks tested for within-group comparisons.

Hypothesis 1: The SLI group will perform below the TLD group on literal and inferential (causal and informational) questions. Figure 1 shows the comprehension question scores, expressed as a percentage of the total possible points, by group and question type in the read-through condition. Figure 2 displays the same results for the think-aloud condition. The figures show that scores were in the expected direction in both conditions: the scores of the SLI group on literal, informational, and causal questions appear lower than the scores of the TLD group. Given, however, the non-normal distribution of the variables, nonparametric statistics were used to compare the data.

A series of Mann-Whitney U tests was conducted to test the hypotheses that the SLI group would perform below the TLD group on literal, informational, and causal questions. The alpha level was adjusted for multiple comparisons: Bonferroni adjusted alpha = .017 (.05/3 comparisons per condition). The results were significant for all three question types in both the read-through and think-aloud conditions. In the read-through condition, the scores of the SLI group were significantly below the TLD scores on literal questions, $U = 16.00$, $z = -3.26$, $p = .001$ (one-tailed); informational questions, $U = 22.50$, $z = -2.87$, $p = .003$ (one-tailed); and causal questions, $U = 8.50$, $z = -3.71$, $p < .01$ (one-tailed). In the think-aloud condition, scores of the SLI group were significantly below scores of the TLD group on literal questions, $U = 26.50$, $z = -2.67$, $p = .007$ (one-

tailed); informational questions, $U = 9.50$, $z = -3.63$, $p < .01$ (one-tailed); and causal questions, $U = 14.50$, $z = -3.36$, $p < .01$ (one-tailed).

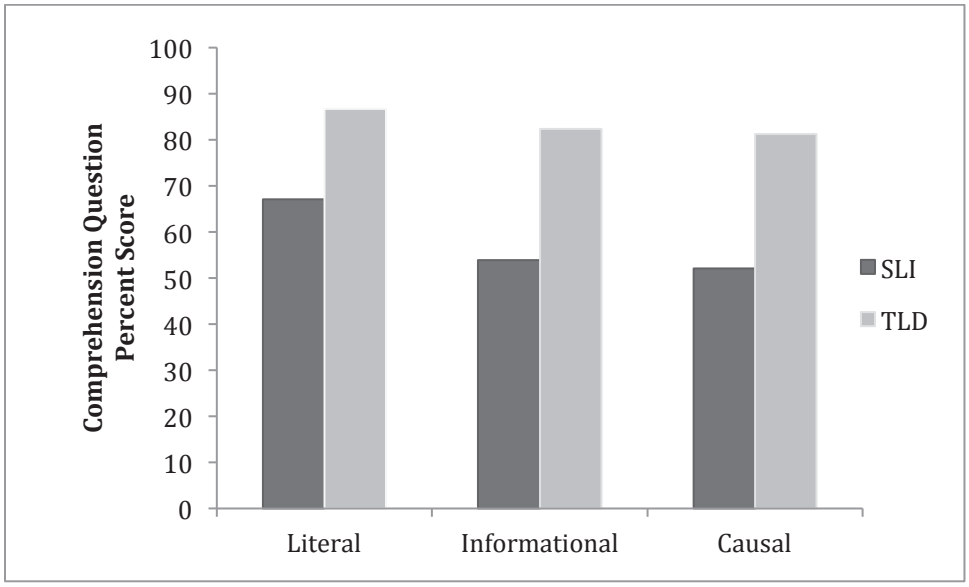


Figure 1. Comprehension question scores (percent of total possible points) for all question types by group in the read-through condition

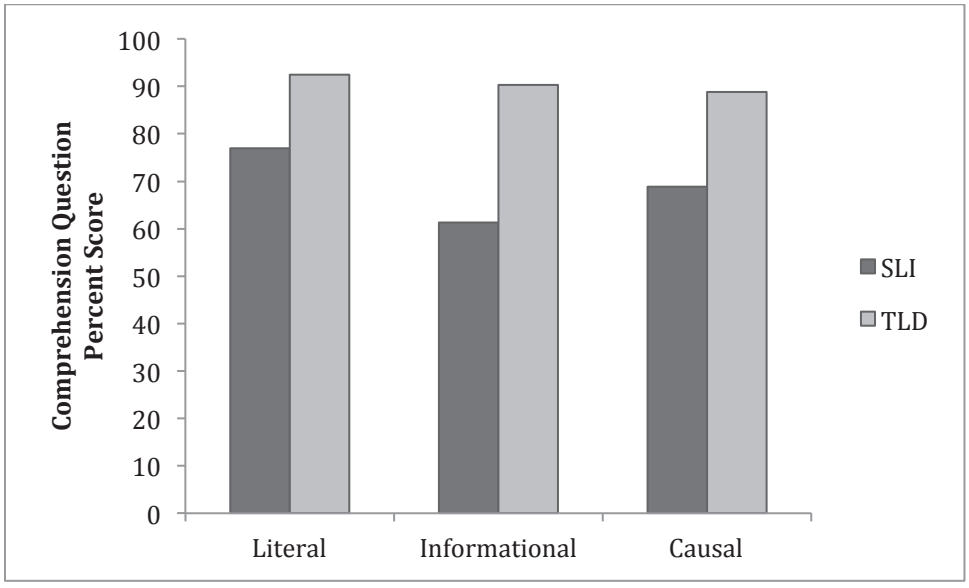


Figure 2. Comprehension question scores (percent of total possible points) for all question types by group in the think-aloud condition

Hypothesis 2: Both groups will have greater difficulty with inferential than

literal questions. Figure 3 displays the literal, informational, causal, and total comprehension scores (expressed as a percentage of the total possible points), by group and by condition. The figure shows that scores were in the expected direction; scores on inferential questions (both informational and causal) appear lower than scores on literal questions.

Wilcoxon Signed Ranks tests were used to test the hypothesis that both groups would have greater difficulty with inferential than literal questions. Scores on literal questions were compared to informational and causal questions. A Bonferroni-adjusted alpha of .017 (.05/3) was again applied. In the read-through condition, the results for the SLI group were significant for the literal-causal comparison, z = -2.19, p = .014 (one-tailed), but not significant for the literal-informational comparison, z = -1.73, p = .042 (one-tailed). For the TLD group, the literal-causal comparison was not significant, z = -.94, p = .17 (one-tailed), nor was the literal-informational comparison, z = -.09, p = .47 (one-tailed). In the think-aloud condition, results for the SLI group were non significant for both comparisons: literal-causal, z = -1.45, p = .073 (one-tailed); and literal-informational, z = -2.04, p = .021 (one-tailed). For the TLD group, results were also non significant: literal-causal, z = -1.17, p = .12 (one-tailed); and literal-informational, z = -.59, p = .26 (one-tailed).

Hypothesis 3: Both groups will have lower scores on questions requiring causal inferences than on questions requiring informational inferences. As can also be seen in Figure 3, the scores on causal questions appear to be slightly lower than scores on informational questions for both groups in the read-through condition, and for the TLD group in the think-aloud condition. This pattern was reversed for the SLI group in the

think-aloud condition, where scores on causal questions appear higher than scores on informational questions.

Wilcoxon Signed Ranks tests (Bonferroni adjustment, $\alpha = .05/3 = .017$) were conducted to test the hypothesis that both groups would have lower scores on questions requiring causal inferences than on those requiring informational inferences. In the read-through condition, the comparison was non significant for both the SLI group, $z = -.13$, $p = .45$ (one-tailed), and the TLD group, $z = -.31$, $p = .38$ (one-tailed). In the think-aloud condition, the results were non significant for both the SLI group, $z = -1.49$, $p = .068$ (one-tailed), and the TLD group, $z = -.62$, $p = .27$ (one-tailed).

Hypothesis 4: Performance of both groups on comprehension questions will be enhanced in the think-aloud condition. Figure 3 shows that scores on all question types, as well as the overall comprehension scores, were in the expected direction for both groups; scores were higher in the think-aloud than in the read-through condition.

Wilcoxon Signed Ranks tests (Bonferroni adjustment $\alpha = .05/4 = .013$) were used to test the hypothesis that the performance of both the SLI and TLD groups would improve significantly in the think-aloud condition. For total comprehension scores, results were nearing significance for the SLI group, $z = -2.19$, $p = .014$ (one-tailed). Results were significant for the TLD group, $z = -2.59$, $p = .005$ (one-tailed).

Comprehension scores were further analyzed by question type to test for within–group differences across condition. For the SLI group, there was no significant difference between conditions for literal, $z = -1.81$, $p = .036$ (one-tailed), or informational $z = 1.10$, $p = .14$ (one-tailed) questions. There was, however, a significant difference for the SLI group on causal comprehension questions across condition, $z = -2.55$, $p = .006$

(one-tailed). For the TLD group, there was no significant difference between conditions for any of the question types: literal, $z = -.95$, $p = .17$, informational, $z = -1.86$, $p = .032$, and causal, $z = -1.06$, $p = .15$.

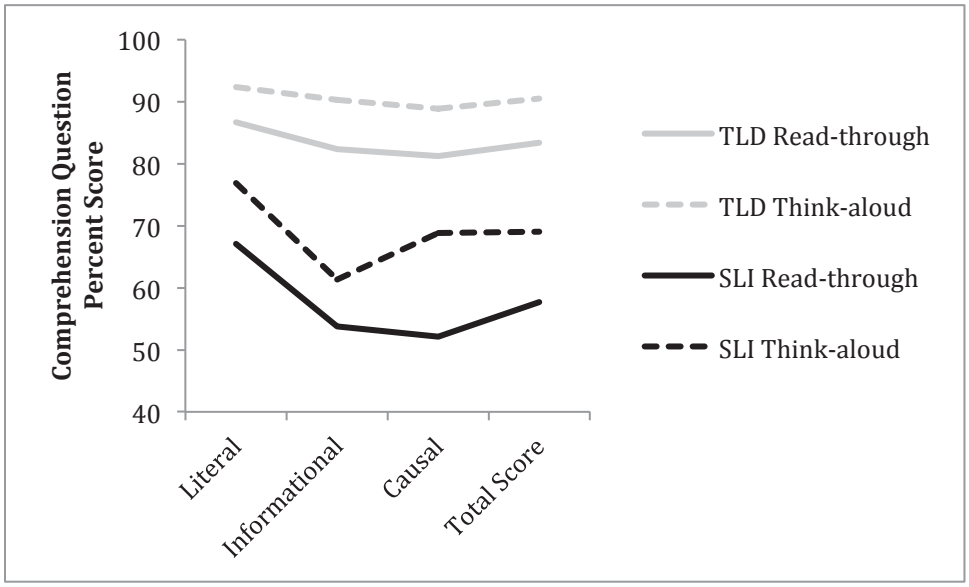


Figure 3. Comprehension question scores (percent of total possible points) for all question types by group in both the read-through and think-aloud conditions

Think-aloud Statements.

Hypotheses 1 & 2: The think-aloud statements of the SLI group will comprise (1) fewer informational inferences and (2) fewer causal inferences than the statements of the TLD group. Given that the think-aloud statements were normally distributed, independent *t*-tests were conducted (Bonferroni adjustment, $\alpha = .05/3 = .017$) to test the hypotheses that the SLI group would make fewer informational and fewer causal inferences than the TLD group. Results were in the expected direction, and were significant. The SLI group made significantly fewer informational inferences, $t(22) = 2.32$, $p = .015$ (one-tailed), and significantly fewer causal inferences, $t(15.2) = 3.09$, $t = .004$ (one-tailed) than the TLD group.

A third independent *t*-test was conducted to explore the possibility that the results for inferences simply reflected a tendency for the SLI group to speak less than the TLD group. The numbers of literal statements made by the SLI and TLD group were compared, under the assumption that differences in volubility, if any, should be reflected in literal statements. The SLI group did not make significantly fewer literal statements, $t(22) = -1.063$, $p = .15$ (one-tailed). These results suggest that the fewer inferences made by the SLI group were not solely the result of the SLI group speaking less than the TLD group.

Hypothesis 3: The SLI group will make less accurate causal inferences than the TLD group in the think-aloud condition. Mann-Whitney U tests (Bonferroni adjustment, $\alpha = .05/2 = .025$) were conducted to test the hypothesis that the SLI group would make less accurate causal inferences than the TLD group, as indicated by a greater proportion of statements scored as 1 or 0 points (considered "inaccurate", or "outlandish", respectively), out of a possible 2 points for each statement. Results were in the expected direction, and were significant for causal inferences scored 1 point, $U = 24.00$, $z = -2.88$, $p = .003$ (one-tailed), and causal inferences scored 0 points, $U = 36.00$, $z = -2.73$, $p = .003$ (one-tailed).

Correlations of causal inferencing. To explore the relationship between language, nonverbal intelligence, working memory, and causal inferencing, Pearson correlations were conducted on composite scores of the CELF-4 (Receptive Language, Expressive Language, And Language Memory), TONI-3 scores, scores on the WMTB-C subtests (Mazes Memory and Block Recall), scores on causal comprehension questions in the read-through and think-aloud conditions, and the total number of causal statements in

the think-aloud. As is shown in Table 2, the CELF-4 composite scores were all significantly and highly correlated with scores on the causal comprehension questions in the read-through and think-aloud conditions. The Expressive Language composite scores of the CELF-4 were significantly and moderately correlated with causal think-aloud statements, and the Receptive Language composite scores were nearing significance (p = .055).

Table 2

Pearson Correlations Among Verbal and Nonverbal Measures, Causal Question Comprehension and Causal Think-aloud Statements

	Block Recall	Mazes Memory	TONI-3	Causal Qs RT	Causal Qs TA	Causal Statements TA
Rec. Language: CELF-4	.61**	.27	.54**	.75**	.61**	.40
Exp. Language: CELF-4	.46*	.56	.43*	.81**	.72**	.48*
Block Recall: WMTB-C	-	.50*	.44*	.40	.11	.49*
Mazes Memory: WMTB-C		-	.35	.06	.07	.14
TONI-3			-	.34	.07	.13
Causal Qs RT				-	.57**	.39
Causal Qs TA					-	.40
Causal Statements TA						-

Note. The CELF-4 composite scores were all highly correlated in the study sample, as they were for the normative sample (Semel et al., 2003); these intercorrelations are not relevant to the present study. Rec = Receptive; Exp = Expressive; Qs = Questions; RT = Read-through; TA = Think-aloud.
* $p < .05$. ** $p < .01$.

Scores on the Block Recall subtest of the WMTB-C were nearing significance with causal comprehension questions in the read-through condition, but were non

significant in the think-aloud condition. The Block Recall scores were significantly correlated with causal think-aloud statements. Scores on the Mazes Memory subtest were not significantly correlated with either the comprehension questions in the read-through and think-aloud conditions, or with causal think-aloud statements, nor were scores on the TONI-3. Scores on causal comprehension questions in the read-through and think-aloud conditions were nearing significance with causal inferences in the think-aloud.

Discussion

The present study was conducted to (1) compare the inferencing abilities of children with SLI to children with TLD, as measured by comprehension questions after reading; (2) to investigate the inferences made during reading by both groups using a think-aloud procedure; and (3) to explore whether use of the think-aloud procedure improves reading comprehension of narrative texts for both groups. The results will be discussed as they relate to theories of reading comprehension, and the results of the think-aloud procedure will be used to shed light on the reading comprehension findings. Limitations, practical implications, and directions for future research will also be discussed.

Reading Comprehension

According to the simple view of reading proposed by Hoover and Gough (1990), reading comprehension is the product of two main components: decoding and oral language comprehension. As each component increases, reading comprehension increases. If either of these two components is impaired, so is reading comprehension. Children with SLI, who, by definition, have difficulties with oral language, are at

increased risk for the development of a reading disability, and often perform below age and grade level on reading comprehension compared to their typically developing peers (Botting & Adams, 2005; Catts et al., 1999; Norbury & Bishop, 2002; Wright & Newhoff, 2001).

Although Hoover and Gough's (1990) model of reading has been proposed as a 'simple' view, the authors argue that reading is indeed a complex activity, and that multiple skills make up both oral language comprehension and decoding, and thus contribute to reading comprehension. One skill, described by van Kleeck (2008) as contributing to reading comprehension via oral language comprehension, is inferencing. Several types of inferences have been described in the literature, some of which readers make to connect parts of text (text-to-text or cohesive inferences) and others which readers make to connect text and world knowledge (text-to-world inferences). Text-to-world inferences, the focus of the present study, include informational and causal inferences. When making informational inferences, readers use their world knowledge to elaborate a single, explicitly stated proposition in the text. In causal inferencing, readers connect explicitly stated propositions in the text by using their world knowledge to fill in information that has only been implied (Van Kleeck, 2008). Causal inferencing has been cited as the most important type of inferencing for successful reading comprehension, as readers who consistently make causal connections by linking story propositions during reading create a coherent mental representation of the story (Trabasso & Magliano, 1996b).

Previous research has shown that children with SLI perform significantly below their typically developing peers on measures of literal and inferential comprehension

(Bishop & Adams, 1990; Norbury & Bishop, 2002; Wright & Newhoff, 2001). The results of the present study are consistent with these findings; children in the SLI group performed significantly below their TLD peers on sets of comprehension questions designed to assess literal and inferential understanding. Given that the participants simultaneously listened to and read the stories during testing, and that the stories were written below grade-level about familiar or accessible topics, the comprehension results are unlikely uniquely due to difficulties with decoding or reading fluency, or to inadequate world knowledge. Indeed, previous research controlling for these variables suggests that difficulties in the reading comprehension of poor comprehenders reflect difficulty with inferencing (Cain et al., 2001).

Consistent with results for poor comprehenders (Laing & Kamhi, 2002), children with SLI in the present study made significantly fewer informational and causal inferences than their TLD peers during the think-aloud. These findings can be attributed to differences in inferencing ability rather than in talkativeness, given that the SLI and TLD groups made a similar number of literal statements on the think-aloud, and produced comparable numbers of independent clauses on a narrative task. In addition to making fewer inferences, the causal inferences made by the SLI group were more likely to be inappropriate or outlandish (given the information provided in the story) than were the inferences made by the TLD group. That is, while children with SLI were able to make inferences, the mental representations of the text they created during reading were likely less coherent and less accurate than the representations of stronger readers. A more impoverished representation could explain poorer performance on the comprehension questions asked after reading.

Inferencing, although described by van Kleeck (2008) as an oral language skill, has alternatively been described in the literature as a skill that relies on both oral language and cognitive skills (Kintsch, 1988; Newton et al., 2010). By definition, children with SLI have oral language difficulties. There is, however, evidence that at least some children labeled as SLI exhibit subtle cognitive difficulties that are not captured by standardized measures (Silliman & Scott, 2006). In the present study, children with SLI performed significantly below their TLD peers on one measure of nonverbal working memory (the Block Recall subtest of the WMTB-C), while group performance was comparable on a second measure from the same battery: the Mazes Memory subtest. The test authors have reported that the two subtests are only weakly correlated (r = .30), suggesting differences in precisely what each of the two tasks measure (Pickering & Gathercole, 2001). Findings for the Block Recall task indicate that for the SLI group, at least some facets of working memory were affected.

Correlational data showed that there were differential associations of the two working memory measures with causal inferencing. Block Recall scores, but not Mazes Memory, were significantly correlated with the language measures, with causal inferencing in the think-aloud procedure, and neared significance with the causal comprehension scores in the read-through condition. Furthermore, the Expressive and Receptive Language composite scores were each significantly correlated with causal comprehension scores in both the read-through and think-aloud conditions. The Expressive Language composite also correlated with causal inferences made during the think-aloud procedure, and the Receptive Language neared significance. Together, these results indicate that nonverbal working memory is related to oral language, both are

related to causal inferencing, and that oral language and working memory contributed to the group differences observed on causal inferencing in the present study.

The correlations observed are in line with Kintsch's (1988) construction-integration model of text comprehension. The model can also be used to explain why the SLI group made significantly more inaccurate and outlandish causal inferences than the TLD group. Kintsch (1988) proposed that language comprehension and working memory, along with world knowledge, are necessary for inference generation and successful text comprehension. According to Kintsch's (1988) model, text comprehension occurs in two phases. In the construction phase, language comprehension contributes to recognition of word meaning and to the generation of inferences. In this phase, difficulties with oral language comprehension experienced by the SLI group in the present study would have led to the generation of incorrect or inaccurate inferences. In the integration phase of comprehension, as each sentence or phrase was being processed and integrated to make sense of the story as a whole, difficulties with working memory would have inhibited the replacement of incorrect or inaccurate inferences with accurate inferences, leading the SLI group to produce lower quality inferences during reading than the TLD group, resulting in inadequate story comprehension.

In terms of performance by inference type, findings by Wright and Newhoff (2001) led to the hypothesis that scores on informational questions would be higher than scores on causal questions for both the SLI and TLD groups. This finding was not replicated in the present study. There was no significant difference between informational and causal scores for either the SLI or the TLD group. This discrepancy can be explained in two ways. First, the stimuli used in the present study differed from

that used in Wright and Newhoff (2001), perhaps enough to elicit varying results. The stories in the present study were intentionally written with numerous gaps in explicit information in order to investigate inferences in the think-aloud condition. The informational inferences required might thus have been more difficult than in the stories used by Wright and Newhoff (2001). Their stories may have had greater coherence as they were used to assess participant's story retelling in addition to their offline inferencing ability. Thus, the informational inferencing invited by the text may have been less challenging, and, in turn, better than the causal inferencing.

Secondly, closer inspection of the test data in Wright and Newhoff (2001) reveals that while expressive language scores were low in the language impaired group, receptive scores were within one standard deviation of the mean. This contrasts sharply with the children in the SLI group in the present study, whose scores on both expressive and receptive language tasks were on average two standard deviations below the test mean. While informational and causal inferences are differentially important to reading comprehension, they could have been equally difficult for children with SLI in the present study due to their mixed expressive and receptive language difficulties and the severity of the impairments, combined with working memory impairments.

Think-Aloud Procedure as a Strategy to Improve Comprehension

Although think-alouds can be considered a "window" into children's thinking, the procedure facilitated comprehension for both the TLD and SLI group in the present study, results that are consistent with previous research using a think-aloud procedure (Laing & Kamhi, 2002; Trabasso & Magliano, 1996a). The overall comprehension of participants in the TLD group improved significantly, and the improvement in the overall

comprehension of the SLI group was borderline significant in the think-aloud condition (p = .014 with a Bonferroni adjusted α = .013). Thinking aloud about each story sentence might have prompted children to make more causal connections while reading than they would normally. This interpretation is supported by findings for children in the SLI group, who improved significantly on causal comprehension across conditions (while scores on literal and informational questions did increase, the improvement was not significant). It appears that when prompted, children with SLI work to create a more coherent mental representation of the text and thus make a greater number of 'critical' causal inferences, but focus less on making 'optional' informational inferences.

Trabasso and Magliano (1996b) proposed that by making an increased number of causal connections during reading, the reader is better able to retain story information in working memory and, in turn, the information is more available when readers answer comprehension questions after reading. This reasoning can be applied to the findings in the present study. The think-aloud condition facilitated 'online' causal inferencing, and presumably decreased the working memory processing load, which could in turn explain the observed improvement in responses to 'offline' causal questions asked after the think-aloud was complete. Indeed, the correlation between the Block Recall working memory task and causal question scores was nearing significance in the read-through condition (p = .055), suggesting that working memory was implicated in answering comprehension questions in that condition. In contrast, working memory was not significantly correlated with the causal question scores in the think-aloud condition (p = .626). These results suggest that the think-aloud procedure might have decreased the working memory processing load required to successfully answer comprehension questions about the story.

Limitations of the Present Study

There are several limitations of the present study that should be addressed. The first relates to the nature and size of the sample. The present study did not include a younger group, matched to the SLI group on some aspect of language ability. In comparing the results of the SLI group to a language-matched group, it might be possible to disentangle the effects of language and working memory on causal comprehension and inferencing. With additional participants, more sophisticated statistical analyses could be employed to determine the contribution of oral language and working memory to inferencing. Furthermore, larger sample sizes would increase statistical power, perhaps allowing several of the study findings, which were borderline or nearing significance, to reach significance.

A further limitation to the present study was in the coding of the think-aloud statements. Interestingly, there was a tendency for the coders to make inferences about the participants' statements, and to subsequently code statements according to *their own inferences*, rather than the child's, making it difficult to obtain intercoder reliability. The 'blindness' of the raters to the study's goals and their more limited knowledge about inferencing seemed to affect their ability to monitor their own thoughts and inferences about the participant's statements, despite extensive training on the coding system and its application. When the thesis supervisor reviewed discrepant coding, her scoring tended to resemble the scoring of the researcher. It thus appears that more in-depth training about inferencing is required in order to accurately score the think-aloud transcriptions. Despite these limitations, the findings of the present study are informative and contribute important information to the existing literature on the inferencing ability and reading

comprehension of school-aged children with language impairments.

Directions for Future Research and Practical Implications

In previous research investigating the reading comprehension of children with language impairments, it has been suggested that the difficulty in reading comprehension observed in children with language impairments is in part due to difficulty with retrieving both the textual and inferential information of the story while simultaneously processing the language of the questions generally used to assess comprehension (Norbury & Bishop, 2002). The think-aloud procedure avoids this difficulty, and could be used to elucidate whether children with language impairments are making specific inferences during reading, but fail to answer comprehension questions requiring the same inferences after reading.

Previous research, as well as the present study, has demonstrated that the think-aloud can be used to identify group differences in children's inferencing ability (Laing & Kamhi, 2002). It could also be used to assess developmental changes in comprehension. There have not yet been any longitudinal studies tracing the development of children's inferencing ability as they learn to read. Although transcribing and scoring is a lengthy process, further research using the think-aloud procedure could nonetheless be informative in identifying developmental changes in the inferencing abilities of typically developing children at various ages and stages of development.

The think-aloud could also be used to assess individual children's reading, or as an intervention strategy. The think-aloud is a practical method that, as has been shown in the present study, facilitates the reading comprehension of children with language and reading impairments. Further research is necessary to determine whether the think-aloud

procedure could be taught to students with language impairments using a direct instruction method, in order to eventually be used as a self-directed strategy to maintain long-term improvements in comprehension. Through direct explanation, teacher modeling, guided practice, application of the particular strategy, and cooperative learning, reading comprehension strategies can be successfully taught within a classroom context (NIFL, 2001). Future research should explore the think-aloud procedure as a strategy to improve the reading comprehension of children with language impairments during shared storybook reading in a classroom setting. It is possible that with effective guidance from a teacher, gains in comprehension could be made for a number of students at once.

Conclusion

The present study has demonstrated that the think-aloud procedure, comprised of general prompts that direct the child's attention and elicit statements that reflect his or her understanding of the text, is effective in improving the reading comprehension of both children with typical language development and children with language impairments. In the present study, the think-aloud procedure revealed that during reading of narrative texts, children with SLI made the same number of literal statements as children in the TLD group, but fewer informational and causal inferences. Additionally, the SLI group made significantly more inaccurate and outlandish causal inferences, leading to a less coherent mental representation of the text and lower comprehension scores. As is explained by Kintsch's (1988) construction-integration model of text processing, both working memory and oral language skills, which were impaired in the SLI group and correlated with causal inferencing, contribute to the generation of inaccurate or

outlandish inferences.

Contrary to previous findings, there was no significant difference between informational and causal comprehension for the SLI group in the present study, suggesting that both types of inferences were equally difficult for children with severe expressive and receptive language difficulties. However, improvement in causal comprehension for the SLI group and overall comprehension for the TLD group in the think-aloud condition suggests that with prompting, readers make more causal inferences, which are critical for comprehension. In doing so, the reader creates a more coherent mental representation of the text, as is reflected in improved performance on offline comprehension questions. The practical implications of these results include the potential of the think-aloud procedure as a strategy to improve the inferential comprehension of children with language impairments, to be implemented by educators and perhaps gradually taught to children so that they might boost their own comprehension, and enjoy the pleasure and learning that comes from adequate text comprehension.

References

Adams, C., Clarke, E., & Haynes, R. (2009). Inference and sentence comprehension in children with specific or pragmatic language impairments. *International Journal of Language & Communication Disorders, 44*(3)*,* 301-318. doi:10.1080/13682820802051788

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29. doi:10.1016/j.jecp.2009.11.003

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry, 31*(7), 1027-1050.

Bishop, D.V.M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech & Hearing Research, 35*(1), 119-129.

Botting, N., & Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *International Journal of Language and Communication Disorders, 40*(1), 49-66. Doi:10.1080/13682820410001723390

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*, 189-201. doi:10.1348/000709904X22674

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence: A Language-Free Measure of Cognitive Ability.* Austin, TX: Pro-Ed.

Cain, K., & Oakhill, J. (1998). Comprehension skill and inference-making ability: Issues

of causality. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling* (pp. 329-342). Mahwah, New Jersey. Lawrence Erlbaum Associates.

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

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850-859.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671-681. doi:10.1037/0022-0663.96.4.671

Catts, H. W., Bridges, M. S., Little, T.D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research, 51*(6), 1569-1579.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142-1157.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331-361.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-

explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145-182. doi:10.1016/0364-0213(89)90002-5

Clarke, P.J., Snowling, M. J., Truelove E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*(8), 1106-1116. doi:10.1177/0956797610375449

Colman, A. (2009). *A dictionary of psychology.* Retrieved from http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e4143

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science, 5*(5)*,* 241-244. doi:10.1111/j.1467-9280.1994.tb00619.x

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*(3), 277-299.

Eickhoff, J., Betz, S.K., & Ristow, J. (2010, June). *Clinical procedures used by speech-language pathologists to diagnose SLI*. Poster presented at the Symposium on Research in Child Language Disorders. Madison, WI.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251. doi: 10.1037/0033-295X.87.3.215

Gillam, S. L., Fargo, J. D., & Robertson, K. S. C. (2009). Comprehension of expository test: Insights gained from think-aloud data. *American Journal of Speech-Language Pathology, 18*, 82-94. doi:10.1044/1058-0360(2008/07-0074)

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10. doi:10.1177/074193258600700104

Gouvernement du Québec (2007). *Organization of educational services for at-risk students and students with handicaps, social maladjustments or learning difficulties.* Retrieved from http://www.mels.gouv.qc.ca/DGFJ/das/orientations/orientations.html

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127-160.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765-778. doi:10.1037/a0015956

Kelso, K., Fletcher, J., & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language and Communication Disorders, 42,* 39-57. doi:10.1080/13682820600693013

Kim, Y-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*(3), 652-667. doi:10.1037/a0019643

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare the inferencing abilities in average and below-average readers. *Journal of Learning*

Disabilities, 6(5), 436-447.

McGhee, R. L., & Lieberman, L. R. (1990). Test-retest reliability of the Test of Nonverbal Intelligence (TONI). *Journal of School Psychology, 28*(4), 351-353. doi:10.1016/0022-4405(90)90024-2

McInnes, A., Humphries, T., Hogg-Johnson, S., & Tannock, R. (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of Abnormal Child Psychology, 31*(4), 427-443. doi:10.1023/A:1023895602957

McKeough, A., Case, R., Bereiter, C., Anderson, V., Adams, M. J., & Hirshberg, J. (1996). *Reproducible Masters: Story Crafting Kindergarten.* Chicago: Open Court.

Montgomery, J. W., Magimairaj, B. M., Finney, M. C. (2010). Working memory and Specific Language Impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology, 19*, 78–94.

National Institute of Child Health & Human Development (2006, August 31st). *Report of the National Reading Panel: Teaching children to read.* Retrieved from http:// www.nichd.nih.gov/publications/nrp/findings.cfm

National Institute of Child Health & Human Development (2010, March 24th). Learning disabilities. Retrieved from http://www.nichd.nih.gov/health/topics/learning_disabilities.cfm

National Institute for Literacy (2001, September). Put reading first: The research building blocks for teaching children to read. Retrieved from

http://www.nationalreadingpanel.org/Publications/publications.htm

Newton, E. J., Roberts, M. J., & Donlan, C. (2010). Deductive reasoning in children with specific language impairment. *British Journal of Developmental Psychology, 28,* 71-87. doi:10.1348/026151009X480185

Norbury, C. F., & Bishop, D. V. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language & Communication Disorders, 37*(3), 227-251. doi:10.1080/13682820210136269

Perfetti, C. A., & Zhang, S. (1996). What it means to learn to read. In M.F. Graves, P. van den Broek, & B. M. Taylor (Eds.), *The first r: every child's right to read* (pp. 37-61). NY: Teachers College Press.

Phillips, L. M. (1988). Young readers' inference strategies in reading comprehension. *Cognition and Instruction, 5*(3), 193-222. doi:10.1207/s1532690xci0503_1

Pickering, S. & Gathercole, S. (2001). *Working Memory Test Battery for Children (WMTB-C).*Toronto: Pearson Assessment Canada.

Rice, M.L. (2009). Children with Specific Language Impairment: Bridging the genetic and developmental perspectives. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 411-431). Malden, MA, USA: Blackwell.

Rice, M. L., Smith, S. D., & Gayán, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment. *Journal of Neurodevelopmental Disorders, 1*(4), 264-282. doi:10.1007/s11689-009-9031-x

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review, 19,* 185-221. doi:10.1007/s10648-006-9024-1

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4).* Toronto: Pearson Assessment Canada. All participants reported speaking English the majority of the time at home with some exposure to a second or third language.

Shanahan, T. (2006). *The National Reading Panel Report: Practical advice for teachers*. Retrieved from http://www.learningpt.org

Silliman, E. R., & Scott, C. M. (2006). Language impairment and reading disability: Connections and complexities, introduction to the special issue. *Learning Disabilities Research & Practice, 21*(1), 1-7. doi:10.1111/j.1540-5826.2006.00202.x

Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal, 55*, 410-413.

Statistics Canada (2010). *Measuring up: Canadian results of the OECD PISA study the performance of Canada's youth in reading, mathematics and science*. Retrieved from http://www.statcan.gc.ca/pub/81-590-x/81-590-x2010001-eng.pdf

Trabasso, T. & Magliano, J. P. (1996a). Conscious understanding during comprehension. *Discourse Processes, 21*, 255-287.

Trabasso, T., & Magliano, J. P. (1996b). How do children understand what they read and what can we do to help them? In M. F. Graves, P. van den Broek, & B. M. Taylor (Eds.), *The first r: Every child's right to read* (pp. 160-188). New York: Teachers

College Press.

Van Kleeck, A. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools, 45*(7), 627-643. doi:10/1002/pits.20314

Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative in adolescent specific language impairment (SLI): A comparison with peers across two different narrative genres. *International Journal of Language & Communication Disorders, 42*(5), 583-605. doi:10.1080/13682820601056228

Windsor, J. (2002). Contrasting general and process-specific slowing in language impairment. *Topics in Language Disorders, 22*(3), 49-61.

Wright, H. H., & Newhoff, M. (2001). Narration abilities of children with language-learning disabilities in response to oral and written stimuli. *American Journal of Speech-Language Pathology, 10,* 308-319. doi:10.1044/1058-0360(2001/027)

Zwiers, J. (2010). Making inferences and predictions. *Building reading comprehension habits in grades 6-12: A toolkit of classroom activities (2nd Ed.),* Newark, DE: International Reading Association.

Appendix A

Experimental Stories, Spache Readability Analyses, and Comprehension Questions

Anna and Her Dog

Anna was walking her dog.

They stopped by the swings to play.

Anna's dog was very good at playing fetch.

Anna threw the ball over and over again.

She threw the ball too far.

It went into the woods.

Anna waited but her dog didn't come back.

Anna looked by the water fountain.

Her dog was not there.

She looked by the baseball field.

The baseball players had not seen her dog.

Anna looked until it started to get dark.

Finally, she started to walk home.

She passed by the swings on her way.

Her dog was waiting for her with the ball.

Anna ran and gave her dog a hug.

(111 words)

Spache Readability Formula for Anna and her Dog =

(0.141 * 6.93 Avg. Number of Words Per Sentence) + (0.086 * 6.30 Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = 2.35 Grade Equivalent

Comprehension Questions

3 Literal:

1) Where did Anna and her dog stop to play with the ball?
2) What game did Anna play with her dog?
3) Where did Anna look for her dog?

3 Causal:

1) Why didn't Anna's dog come back?
2) Why did Anna's dog wait for her by the swings?
3) How did Anna feel when she found her dog?

3 Informational:

1) Where did Anna take her dog for a walk?
2) Who did Anna ask if they had seen her dog?
3) What time of day was it when Anna's dog got lost?

Andrew Goes Skating

Andrew was skating on a small pond.

He was skating very fast.

He was wearing his wool hat, his gloves, and his scarf.

He skated to the middle of the pond.

The ice was thin there.

The ice started to crack.

Andrew cried out when the ice broke.

He crashed through the ice.

A man was walking by the pond.

He rushed out to save Andrew.

The man grabbed Andrew by both arms.

Andrew's clothes were all wet.

When he got home, Andrew got wrapped up in a warm blanket.

He sat down by the fire.

He drank a cup of hot chocolate.

He smiled and gave his mom a hug.

(111 words)

Spache Readability Formula for Andrew Goes Skating =

(0.141 * *6.93* Avg. Number of Words Per Sentence)+ (0.086 * *10.81* Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = *2.74* Grade Equivalent

Comprehension Questions

2 Literal:

1) What was Andrew doing at the start of the story?
2) Who rescued Andrew?

3 Causal:

1) Why did Andrew go where the ice was thin?
2) Why did Andrew sit by the fire when he got home?
3) How did Andrew feel when he got home?

4 Informational:

1) What was Andrew wearing on his feet to keep warm?
2) What season was it?
3) Where did Andrew fall through the ice?
4) What did Andrew do with his wet clothes when he got home?

Mark gives a Gift

Mark wanted to give his mother a gift.

He wanted it to be very special.

Mark thought for a long time.

He knew that chocolate cake was his mother's favorite.

Mark found all of the ingredients.

He followed the instructions.

First he mixed in the sugar, butter and eggs.

He poured the batter into the pan.

He put the cake in the oven.

Mark took the cake out when it was done.

Mark dropped the pan.

The cake was okay, but it had a crack in the middle.

Mark had an idea.

He made some icing.

He put lots of icing on the cake.

He lit the candles.

Mark’s mother was very happy.

(112 words)

Spache Readability Formula for Mark Gives a Gift =

(0.141 * *6.64* Avg. Number of Words Per Sentence)+ (0.086 * *9.73* Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = *2.61* Grade Equivalent

Comprehension Questions

3 Literal:

1) What was Mark's mother's favourite dessert?
2) What did Mark put in the cake batter?
3) Where did Mark put the cake?

3 Causal:

1) Why did Mark want to give his mother a gift?
2) Why did Mark drop the pan?
3) Why did Mark put lots of icing on the cake?

3 Informational:

1) Where did Mark get the instructions to make the cake?
2) Where did Mark find all of the ingredients?
3) What else did Mark mix into the batter?

Gabby and the Stroller

Gabby was walking down the street.

She turned around the corner.

In the alley there was a big pile of garbage.

Gabby saw an old stroller.

She ran to find her friends.

They were playing in the park.

Come and see what I found, she said.

Gabby and her friends got the stroller down.

They were careful not to get hurt.

They took the top part off of the stroller.

They hammered an old piece of wood onto the wheels.

They made a steering wheel.

The children went back to the park.

Gabby got into the go-cart and raced down the path.

The crowd cheered when she won.

(108 words)

Spache Readability Formula for Gabby and the Stroller =

(0.141 * 7.2 Avg. Number of Words Per Sentence)+ (0.086 * 12.96 Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = 2.96 Grade Equivalent

Comprehension Questions

3 Literal:

1) Who did Gabby go to find?
2) What did the children hammer onto the wheels?
3) Where did the children go when they had finished making the go-cart?

3 Causal:

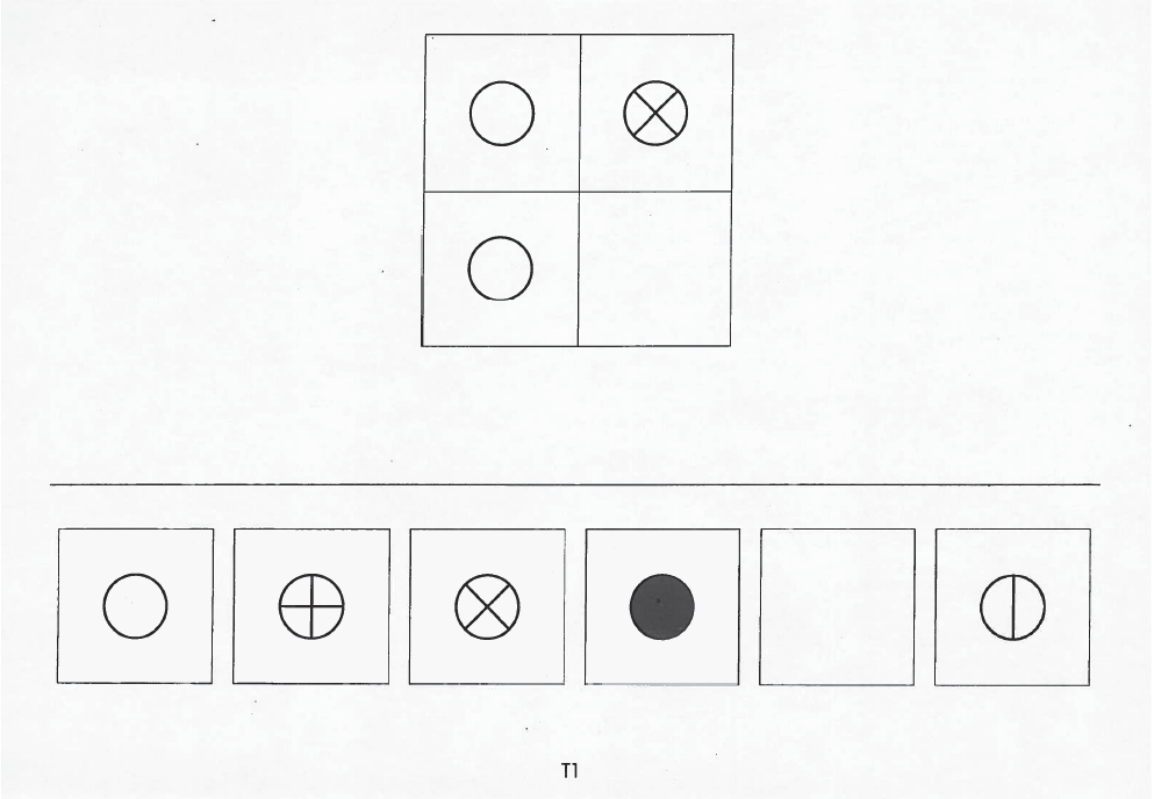
1) Why did Gabby tell her friends about the stroller?
2) Why did the children go back to the park?
3) How did the children feel when they saw Gabby racing in the go-cart?

3 Informational:

1) Where did Gabby find the old stroller?
2) What was the stroller doing on top of the pile of garbage? **(Omitted from Analyses)**
3) What was in the pile of garbage that could have hurt the children?

Appendix B

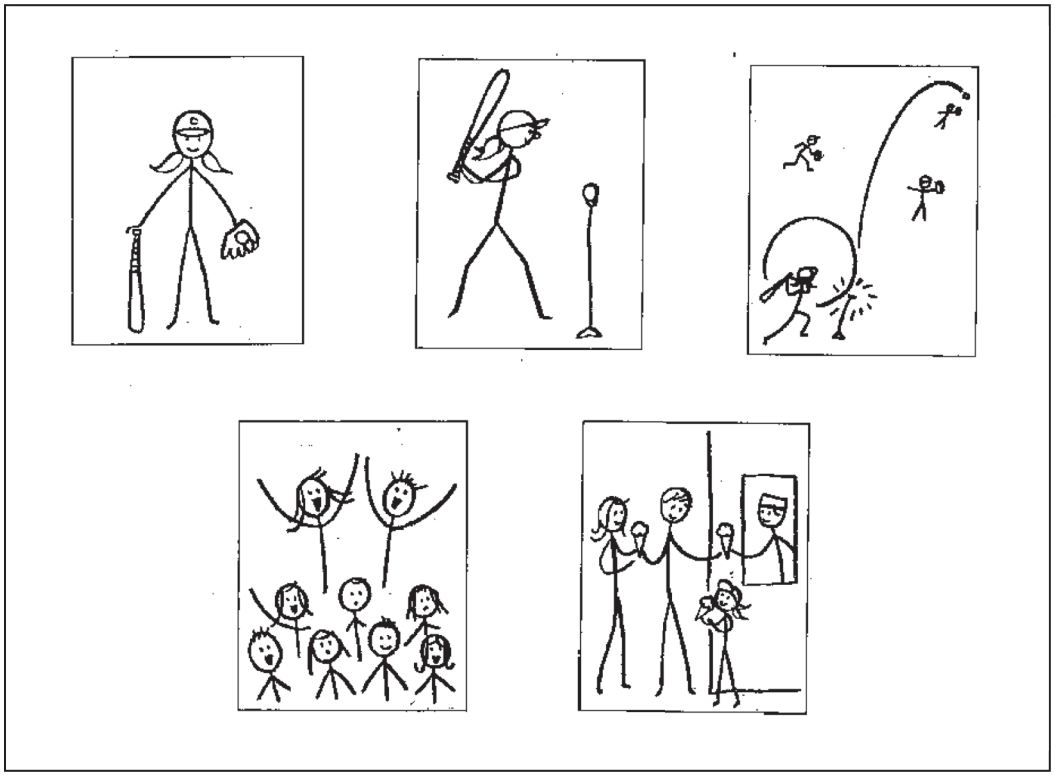
TONI-3 Sample (Training) Item
Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997).

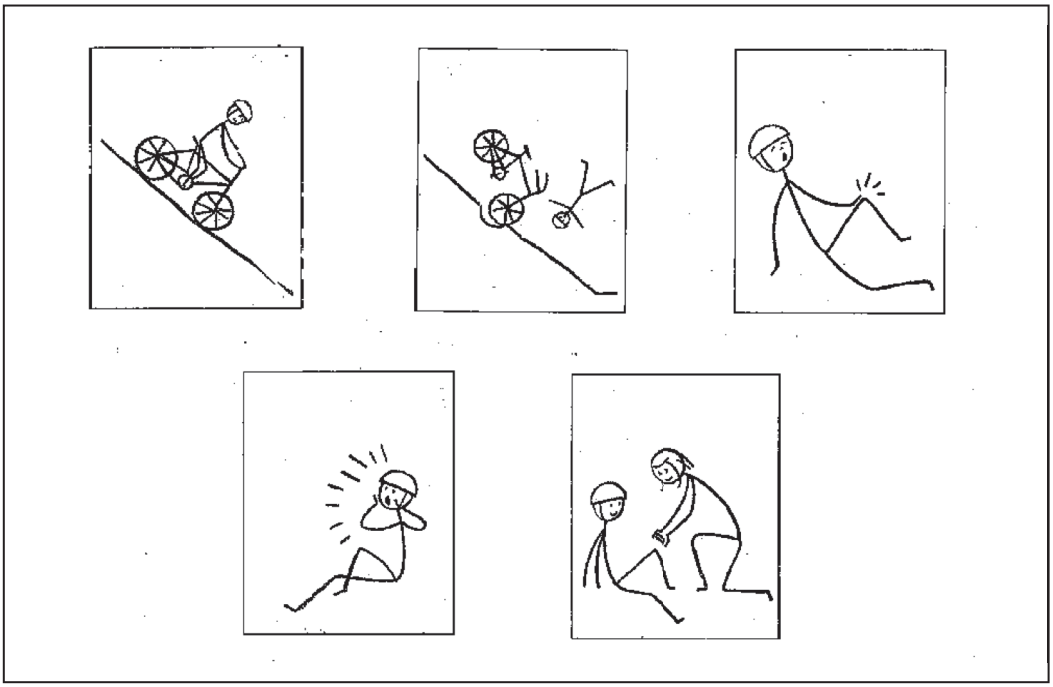


Appendix C

Picture Stories 1 and 2

McKeough, A., Case, R., Bereiter, C., Anderson, V., Adams, M. J., & Hirshberg, J. (1996).





Appendix D

Think-aloud Demonstration Stories and Spache Readability Analyses

Ivan and the Dragon

Prince Ivan lived in a castle.

He had nobody to play with.

There was a big dragon that lived in the forest nearby.

One day, the prince went for a walk.

Ivan heard somebody crying in the forest.

The dragon was lonely too.

Ivan was happy to have found a friend.

(51 words)

Spache Readability Formula for Ivan and the Dragon =

(0.141 * *7.28* Avg. Number of Words Per Sentence) + (0.086 * *13.72* Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = *3.04* Grade Equivalent

Rachel and the Giant

(Adapted from Trabasso & Magliano, 1996a)

Rachel was a great fighter.

One day, Rachel heard that a giant was hurting people in her town.

Rachel decided to scare the giant away.

Rachel waited until dark.

When the giant came, Rachel shot an arrow at him.

Rachel hit him and the giant fell down.

All the people in the town were happy.

(55 words)

Spache Readability Formula for Rachel and the Giant =

(0.141 * *7.85* Avg. Number of Words Per Sentence) + (0.086 * *1.81* Percent of Words in Sample Not Found on Spache Revised Word List) + 0.839 = *2.10* Grade Equivalent

Appendix E

Think-aloud Demonstration Story Scripts

Ivan and the Dragon

Prince Ivan lived in a castle.

Castles are usually really big and made of stone. Because he lives in a big castle, he probably has a big family too. Maybe the Queen and King and his brothers and sisters live with him there.

He had nobody to play with.

So Prince Ivan is probably really lonely. Maybe he doesn't have brothers and sisters after all.

There was a big dragon that lived in the forest nearby.

He's probably a scary dragon. Maybe he lives in a cave in the forest. People probably stay out of the forest because the dragon is there.

One day, the prince went for a walk.

Maybe Ivan goes on lots of walks because he doesn't have any friends, and it's something for him to do. It's probably a nice, warm day out, and it would be a nice day for a walk.

Ivan heard somebody crying in the forest.

He thought he was all alone! He will probably go to see who is crying. But that means he has to go into the forest, and it might be dangerous because of the dragon.

The dragon was lonely too.

It looks like it was the dragon who was crying! Maybe he's not so scary after all, but he has no friends because people think he is scary, just like I did.

Ivan was happy to have found a friend.

So it looks like Ivan and the dragon will become friends. They were both really lonely! This is very nice for both of them.

Rachel and the Giant

Rachel was a great fighter.

So, Rachel was very good at fighting. She probably fought lots of fights and maybe she won a lot.

One day, Rachel heard that a giant was hurting people in her town.

This probably made Rachel really mad. Because she's a good fighter, she will probably want to fight the giant.

Rachel decided to scare the giant away.

Since Rachel's a great fighter, she's probably really brave. She doesn't want the giant to keep hurting people. Maybe she will fight the giant so that he will stop hurting people.

Rachel waited until dark.

So it looks like Rachel has a plan. Maybe she waited until dark because she wants to sneak up on the giant and she doesn't want the giant to see her coming.

When the giant came, Rachel shot an arrow at him.

Since Rachel's such a great fighter, she probably had very good aim even though it was dark. Hopefully she'll hit the giant.

Rachel hit him and the giant fell down.

Oh! She did hit him. The arrow probably hurt the giant. Hopefully the giant will leave and the people in the town don't need to worry about getting hurt anymore.

All the people in the town were happy.

So everybody in the town is happy because Rachel was such a great fighter, and they don't need to worry anymore about the big giant coming to hurt them or bother them. Rachel is probably really happy and relieved too!

Appendix F

Story Stimulus Book Sample Pages

(Think-aloud then Read-through sample pages)

Anna was walking her dog.

Anna was walking her dog.

Appendix G

Assent Script and Story Prompts

Assent Script
I'm here to learn about how people understand stories. Today we are going to be reading stories, answering some questions about the stories, and doing some other activities that will help me see how people understand what they read. Some of these activities might be easy, and some might be hard. If ever you find that the activities are too hard and you want to take a break, you can let me know. If you ever decide that you want to stop doing the activities altogether, you can let me know that too, and that's fine. We will just go back to class.

Picture Story Prompt
Now we are going to warm up our voices a little. Can you tell me as much as you can about the story you see here?

Read-through Prompt
Now we are going to read a story. We will listen to the story out loud, and we will follow along in our heads. At the end of the story, I will ask you some questions about it, so do your best to understand the story as we read.

Think-aloud Prompt
Now we are going to read a story. This time we will do something a little different. We will listen to one sentence of the story at a time, and we will follow along in our heads. You will see a picture like this on the page *(point to Think-aloud picture prompt)*. When I point to this picture, you will explain to me what you understand about the story. I will show you what we are going to do.

Give demonstration.

Now it is your turn. Just like when I did it, we will listen to one sentence of a story at a time, and we will follow along together. When I point to the picture, this time *YOU* will tell me what you understand about the story. At the end of the story, I will ask you some questions about the story, so do your best to understand. Ready?

Comprehension Question Prompt
Now I am going to ask you some questions about the story. It's okay to guess at the answers.

If child gives a one point/incomplete answer, say: Can you tell me a little more?
If child gives no response or says "I don't know", say: Do you want to guess?

Appendix H

Experimental Story Scoring Keys

Anna and her Dog

	P=Prompt (Can you tell me more/Do you want to guess?) **If prompting elicits a response with a lower score, then only count response given before the prompt.**
1.	**Where did Anna take her dog for a walk?** *2 points:* The park *1 point:* by the swings *0 points:* unrelated/incorrect (the water fountain/baseball field)/no response
2.	**Where did Anna and her dog stop to play with the ball?** *2 points:* by the swings *1 point:* park/outside/field/close to the woods *0 points:* water fountain/baseball field/unrelated/no response
3.	**What game did Anna play with her dog?** *2 points:* fetch/throws the ball and the dog goes and gets it *1 point:* ball/catch *0 points:* unrelated/incorrect/no response/baseball
4.	**Why didn't Anna's dog come back?** *2 points:* it couldn't find the ball in the woods/it got lost/any reason for distraction (e.g., Saw squirrel/another dog) *1 point:* she threw the ball too far (with no explanation), 'he was looking for it' with no explicit mention of ball/went looking and when he came back Anna wasn't there *0 points:* unrelated/no response
5.	**Where did Anna look for her dog?** *2 points:* by the water fountain + at the baseball field (optional- and by swings) *1 point:* one of two elements, or one of two with an incorrect answer (and by the swings)/at the park/everywhere *0 points:* incorrect (in the woods)/no response
6.	**Who did Anna ask if they had seen her dog?** *2 points:* the baseball players *1 point:* unspecific (people in the park) *0 points:* incorrect/no response
7.	**What time of day was it when Anna's dog got lost?** *2 points:* afternoon/daytime/indication of it being before dark *1 point:* it was getting dark *0 points:* night/dark/unrelated/no response
8.	**Why did Anna's dog wait for her at the swings?** *2 points:* It's where they were playing/where they started/so Anna could find the dog/he came back while Anna was looking *1 point:* it wanted to play with the ball (no mention of returning)/he didn't know where Anna was *0 points:* incorrect/unrelated/no response
9.	**How did Anna feel when she found her dog?** *2 points:* positive emotion + reason/elaboration (e.g., she thought she'd lost the dog forever, she ran and gave the dog a hug) *1 point:* positive emotion but with no reason or elaboration/reason but with a wrong emotion (e.g., She was worried and sad and giving the dog hugs) *0 points:* reason but with no emotion/incorrect/unrelated/no response

Andrew Goes Skating

	P=Prompt (Can you tell me more/Do you want to guess?) **If prompting elicits a response with a lower score, then only count response given before the prompt.**
1.	**What was Andrew doing at the start of the story?** *2 points:* Skating + On the pond *1 point:* getting ready to go skating/just skating/wanted to skate *0 points:* Incorrect (e.g., Walking by a pond)/No response
2.	**What was Andrew wearing on his feet to keep warm?** *2 points:* Wool/warm socks/socks /socks + skates/skates with fur in them *1 point:* skates/socks + incorrect footwear (e.g., Socks & boots) *0 points:* boots/Incorrect/No response
3.	**What season was it?** *2 points:* Winter/late fall or early spring with mention of ice breaking *1 point:* Beginning of spring (without mention of ice breaking) *0 points:* Incorrect/No response
4.	**Where did Andrew fall through the ice?** *2 points:* In the middle (of the pond) *1 point:* On the pond (no mention of middle) *0 points:* Incorrect (e.g., Pond + wrong location)/No response
5.	**Why did Andrew go where the ice was thin?** *2 points:* He didn't know (it was thin)/He was skating and didn't realize *1 point:* he wanted to go to the middle/was skating fast (no mention of not knowing ice was thin)/he was in a race *0 points:* No response/unrelated/incorrect
6.	**Who rescued Andrew?** *2 points:* The man/a guy + who was walking by the pond *1 point:* a man/Andrew's dad (no mention of walking by pond) *0 points:* Incorrect/No response
7.	**What did Andrew do with his wet clothes when he got home?** *2 points:* He took them off/He hung them up/dried them (in front of the fire) *1 point:* He threw them out/ told his mom (without mention of mom doing something) *0 points:* Incorrect/No response
8.	**Why did Andrew sit by the fire when he got home?** *2 points:* To get warm/he was cold (from falling in pond)/to get dry *1 point:* because he fell through the ice (no mention of being cold/getting warm) *0 points:* Incorrect/Unrelated
9.	**How did Andrew feel when he got home?** *2 points:* Positive emotion + reason/elaboration (e.g., Happy/relieved + to be home/he was safe/gave his mom a hug/mom was taking care of him) *1 point:* emotion without reason/better with no mention of falling through ice/cold/neg. emotion with reason (Sad because his clothes were wet) *0 points:* Incorrect/No response/reason without mention of any emotion

Mark Gives a Gift

	P=Prompt (Can you tell me more/Do you want to guess?) **If prompting elicits a response with a lower score, then only count response given before the prompt.**
1.	**Why did Mark want to give his mother a gift?** *2 points:* it was her birthday *1 point:* because he wanted to make her happy/do something nice for her/because he loved her/to surprise her/he liked giving her presents/mother's day *0 points:* Unrelated/Incorrect/No response
2.	**What was Mark's mother's favourite dessert?** *2 points:* chocolate + cake *1 point:* one of two elements *0 points:* Unrelated/Incorrect/No response
3.	**Where did Mark get the instructions to make the cake?** *2 points:* a recipe (book)/on the box/ from a book *1 point:* he already knew how to make it/from the store *0 points:* Unrelated/Incorrect/No response
4.	**Where did Mark find all of the ingredients?** *2 points:* the kitchen/the cupboards/he went to the store/fridge *1 point:* His mom gave them to him/around the house *0 points:* Unrelated/Incorrect (e.g., cookbook) /No response
5.	**What did Mark mix in the cake batter?** *2 points:* sugar + butter + eggs (credit if all three elements in responses to Q5&Q6) *1 point:* response with one missing element *0 points:* Unrelated/Incorrect (e.g. icing)/No response/only 1 of 3 elements
6.	**What else did Mark mix into the batter?** *2 points:* cocoa or chocolate (+ other cake ingredient) (Credit if mentioned in Q5) *1 point:* any baking ingredient not mentioned in story (e.g. water/flour) with no mention of chocolate or cocoa *0 points:* Unrelated/Incorrect (e.g., icing)/No response
7.	**Where did Mark put the cake?** *2 points:* in the oven *1 point:* in the pan *0 points:* Unrelated/Incorrect (e.g., on the counter)/No response
8.	**Why did Mark drop the pan?** *2 points:* it was hot/he was not wearing oven mitts *1 point:* it was an accident/heavy/he was clumsy/it was slippery/hands were slippery *0 points:* Unrelated/Incorrect/No response
9.	**Why did Mark put lots of icing on the cake?** *2 points:* to hide or cover the crack/make the cake look nice because there was a crack *1 point:* there was a crack (no mention of hiding/covering the crack), so it would look nice *0 points:* Unrelated/Incorrect (e.g., so it would taste good) /No response

Gabby and the Stroller

	P=Prompt (Can you tell me more/Do you want to guess?) **If prompting elicits a response with a lower score, then only count response given before the prompt.**
1.	**Where did Gabby find the old stroller?** *2 points:* In the alley + On top of/in the pile of garbage *1 point:* 1 of 2 elements, near garbage *0 points:* Incorrect/No response
2.	**What was the stroller doing on top of the pile of garbage?** *2 points:* Somebody threw it away/It was garbage/broken *1 point:* Sitting on top of the pile (no indication of being thrown away/garbage) *0 points:* Incorrect/No response
3.	**Who did Gabby go to find?** *2 points:* Her friends + In the park *1 point:* No mention of park, or inaccurate (e.g., the kids, a friend, the boy/girl) *0 points:* Incorrect/No response
4.	**Why did Gabby tell her friends about the stroller?** *2 points:* Help getting it down/wanted to turn it into a go-cart/use it as a go-cart *1 point*: to race it/wanted to show it to them/fix it with no mention of turning into go-cart *0 points:* Incorrect/No response
5.	**What was in the pile of garbage that could have hurt the children?** *2 points:* any sharp object (e.g., (broken) glass or rusty nails)/mention of object + could have tripped on/something that could have fallen on them/ "something sharp" with mention of how could have hurt children *1point:* Mention of any object that might be found in the garbage with no indication or unclear how it could hurt the children/"something sharp" (unspecific)/garbage was big and could have fallen *0 points:* Unrelated/No response
6.	**What did the children hammer onto the wheels?** *2 points:* old piece of wood/wood *1 point:* steering wheel *0 points:* Incorrect/No response
7.	**Where did the children go when they had finished making the go-cart?** *2 points:* back to the park/to the race *1 point:* path, with no mention of park *0 points:* Incorrect/no location (e.g., they raced)/No response
8.	**Why did the children go back to the park?** *2 points:* mention of go-cart & race *1 point:* To play (on the go-cart)/ride the go-cart with no mention of the race, or race with no mention of go-cart *0 points:* Incorrect (e.g., to play on the swings)/No response
9.	**How did the children feel when they saw Gabby racing in the go-cart?** *2 points:* Positive emotion + reason/elaboration (e.g., Happy/proud + she won, the go-cart they built was fast, they cheered) *1 point:* Any positive emotion with no reason *0 points:* a negative emotion/unrelated/incorrect/No response

Appendix I

Sample Think-Aloud Transcription

	Mark Gives a Gift Participant: C05MQ Date of birth: Session: 2-1 Date: Condition: TA					
	Think-Aloud Statements	**Caus.**	**Info**	**Eval**	**Lit.**	**Un.**
1.	**Mark wanted to give his mother a gift.**					
	1. He probably wanted to give his mother a gift					
	2. Because it might be mother's day					
2.	**He wanted it to be very special.**					
	1. He really liked his mom					
	2. So he was making it very special					
3.	**Mark thought for a long time.**					
	1. Mark was thinking					
	2. Because he probably couldn't think of something really cool					
4.	**He knew that chocolate cake was his mother's favorite.**					
	1. I think he will decide to make a chocolate cake					
5.	**Mark found all of the ingredients.**					
	1. He probably helped his mom whenever she made chocolate cake					
	2. So maybe he might know how to make it					
6.	**He followed the instructions.**					
	1. His mom probably had a cake recipe book					
7.	**First he mixed in the sugar, butter and eggs.**					
	1. So I guess he has a recipe book					
8.	**He poured the batter into the pan.**					
	1. His mom probably had batter in the cupboards					
9.	**He put the cake in the oven.**					
	1. He should be careful					
	2. Because he might burn himself					
10.	**Mark took the cake out when it was done.**					

	1. So I guess he didn't burn himself					
	2. And he did it easily					
11.	**Mark dropped the pan.**					
	1. His hands were probably slippery					
	2. And the tray fell out of his hand					
12.	**The cake was okay, but it had a crack in the middle.**					
	1. It probably because he dropped it					
	2. And the tray landed right on the floor					
13.	**Mark had an idea.**					
	1. I think he might take some icing					
	2. And glue it together					
14.	**He made some icing.**					
	1. So he probably made some					
	2. Instead of getting some					
15.	**He put lots of icing on the cake.**					
	1. I think there might be too much icing					
	2. And the cake might fall apart again					
16.	**He lit the candles.**					
	1. I guess the cake didn't fall apart					
	2. And he lit the candles					
	3. And it was his mom's birthday					
17.	**Mark's mother was very happy.**					
	1. So it didn't go after a disaster after all					
	2. So I think it was his mom's birthday and stuff					
	Total Statements					
	Total Statements Scored 2 points					
	Total Statements Scored 1 point					
	Total Statements Scored 0 points					
	Proportion of Total Statements					

Statement Type	2 points	1 point	0 points
Literal			
Informational			
Causal			
Inferential			
Prop. Total Statements Inferential			

Appendix J

Transcription Segmenting Rules

Rule	Example
Segment all utterances into separate clauses, which communicate **a main idea.** **Generally,** each clause will be indicated by a main verb along with any of the following rules:	○ "Peter **went** to the store." "The dog **ran** after the ball."
Include in a **single clause:** • Infinitives ○ Simple (including *wanna*/*gonna* etc) ○ 'wh' (marked by conjunctions *what, who, where, when* **and** by *to)* ○ unmarked (headed by *make, help, watch,* or *let,* with no *to* marker) • Full propositional complements expressing actions or events in the story (headed by 'cognitive' verbs e.g., *think, guess, wish, know, hope, wonder*; may/may not contain the conjunction *that*; or headed by 'speech' verbs i.e., *said, shouted, told, asked* (introducing) dialogue • Clauses expressing participant's perspective (e.g., *I think … I know… I forgot … I assume… I guess … This means…It seems…* etc.) • Relative clauses (i.e., clauses with a relative pronoun such as 'that' or 'which') • Evaluation: reader makes a judgement/evaluation of an action or event described by focal sentence	○ "He has to move." "They wanna run." ○ "I know what to do." ○ "Watch me jump." "Help me pick these up." ○ "She wondered *what* she would do next." ○ "She said *he's a scaredy-cat*." ○ "He thought that the boy was drowning." ○ "*I assume that* he went into the cave." ○ "*I think that* it's his mother's birthday." ○ "That's not the kind *that* I like" ○ "She wondered which she should pick." ○ "Oh no!" /"That's good news!"/ "I'm glad (that) he was careful."
Include as **separate clauses**: • Infinitives expressing a separate idea	○ "They ran down the street/ to catch

or answer *"why?"*	the ball." ○ "It was too heavy/ so we needed to get more people/ and lift it together."
• *Wh* Clause (e.g., *what, who, where, when, how, why*), or temporal terms (e.g., while, before, after, once)	○ The ice broke /*when* he skated to the middle."
• Conditional clauses	○ "If he doesn't put the icing on/ the crack will show. "
• Conjoined clauses (using *and, or, so, but*), that express new ideas, even if subject and/ or verb is not restated	○ "He turned on the computer / and sat in the chair" ○ "He was relieved/ and happy."
• Subordinate clauses	○ "The cat ran up the tree / because it was scared of the dog"
• Implied verbs: if verb is absent but implied (ellipsis) but there is more than one idea that is expressed, code each idea as a separate clause.	○ "He wants to make a cake/ A very special one."

Appendix K

Think-aloud Statement Scoring Key

Number statements sequentially by focal sentence: 1.1, 1.2 … 6.1, 6.2 etc.

Chain: a series of explanatory or informational statements. One statement may become the basis of another statement, either literal or inferential.

Decision Sequence

Consider each statement in order from the first statement to the last statement in a chain. Always try to relate a statement 1) to the focal sentence 2) to previous statements in the chain, beginning with the first statement and moving down the chain.

1. IS THE STATEMENT LITERAL? → LITERAL
   Is the statement a repetition or paraphrase of the current focal sentence or a statement in the current chain?
   e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
   Current TA Statement: He fell off of his bike.
   - Inserting character names for he/she (or vice versa) is not considered as added information (therefore typically considered as literal)
   - Watch out for synonyms and fillers including 'because' & 'so' with no causal link
   - If a statement is a literal repetition of the current focal sentence but is not the first statement following the focal sentence or is a literal repetition of a previous focal sentence AND the statement corresponds to causal (as in #2), score as causal
2. DOES THE STATEMENT COMMUNICATE CAUSE AND EFFECT? → CAUSAL
   - A child might say "because" or "so" before every statement, this does not always indicate causality
   - The statement or focal sentence answers a *'why'* question

   Does the current statement answer a why question/express the cause about the focal sentence/statement in a chain preceding it?
   e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
   Current TA Statement: He rode his bike over a rock.
   a. Cause: Does the current think-aloud statement answer a 'why' question related to the immediate focal sentence (Read <u>focal sentence</u>, ask 'why': If current think-aloud statement answers, code as causal)?
   b. Cause: Does the current think-aloud statement answer a 'why' question related to a statement in the immediate chain (Read first or previous <u>think-aloud statement</u>, ask 'why': If current think-aloud statement answers, code as causal, continue down chain)?

Does the statement express the effect of events/actions/states, etc. stated either in the focal sentence or own TA statement?

e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
Current TA Statement: He scraped up his knee.

c. Effect: Does the focal sentence answer a 'why' question related to the current statement (Read current think-aloud statement, ask 'why'; if focal sentence answers, code as causal)?
d. Effect: Does a previous statement in the immediate chain answer a 'why' question related to the current statement (Read current think-aloud statement, ask 'why'; if previous think-aloud statement answers, code as causal)?

- Remember: If a statement is a literal repetition of the current focal sentence but is not the first statement following the focal sentence or of a previous focal sentence AND the statement corresponds to causal (as in #2), score as causal.
    - In other words: Does the speaker seem to be repeating a previous focal sentence in order to tie events together?
- Verify cause and effect *with "If not A, then not B",* where either A or B can be the focal sentence or a statement in a chain

OR

DOES THE STATEMENT INVOLVE PLANS, INTENTIONS OR EXPLICIT REFERENCE TO **FUTURE EVENTS** → CAUSAL

e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
Current TA Statement: He's going to put a band-aid on his knee.

- *"If-Then"* statements are typically scored as causal, but may refer to already stated events/states and thus be coded as other items.
    a. Does it answer a 'what will/will not/might/might not happen next' question related to the focal sentence?
    b. Does it answer a 'what will/might happen next' question related to a statement in a chain?
    c. Is the speaker referring explicitly to a preceding focal sentence?

3. DOES THE STATEMENT FIT THE INFORMATIONAL CATEGORY? → INFORMATIONAL
    - Is the current statement a literal repetition or paraphrase of the current focal sentence with added information?
    - Remember: Inserting character names for he/she is not considered as added information (therefore typically considered as literal)

    e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
    Current TA statement: when he was riding down the hill.

    a. Does it answer a 'what, who, where, when' question related to the focal sentence or to characters, events, or objects more generally?
    b. Does it answer a 'what, how, who, where, when' question related to a statement in a chain or to characters, events, or objects more generally?
4. IS THE STATEMENT EVALUATIVE? DOES IT ELABORATE A LITERAL STATEMENT WITH AN EVALUATION?
    - Embedded evaluations (e.g., nice cup of cocoa, daring boy) are typically coded as informational

- Ignore comments where the speaker evaluates his/her own think-aloud statements e.g., "I was wrong" or "I don't know"

e.g., Focal sentence/Previous TA statement: The boy fell off his bike.
Current TA statement: That's really scary.

a. Is the statement marked by the use of "that/it is _____"? e.g., "That's bad!"/"Oh no, that's dangerous!"
b. Does the reader judge/evaluate a state, event, action, or object described by the focal sentence/preceding statement based on world knowledge, social conventions, or own values?

Scoring

Literal

- 2 points: close paraphrase (i.e., pronoun/noun exchange) or exact repetition of the sentence read, no extra information given

  - Exact or near-exact repetition of focal sentence or part of focal sentence that maintains important information (e.g., Changing a→ the, changing subject to pronoun)
  - Close paraphrase if the clause represents a transformation of the focal sentence or of a previous statement (such as changing tense) that preserves its meaning

- 1 point: partially correct but makes an error that leads to a significant change in meaning or renders part of a sentence ambiguous. Points should not be deducted for grammatical errors that do not change meaning.
- 0 points: statements that are incorrect, don't make sense

Causal, Informational & Evaluative Inferential Statements

- 2 points: relevant and logically linked to the text or to a previous inference and makes sense given the context of the story
- 1 point: inference related to the story, but information is omitted or wording is such that the inference is not fully accurate
- 0 points: "outlandish", inference doesn't make sense given the information provided in the story. e.g., "Gabby went back to the garbage factory"

Working Definitions

Causal

- Statement provides an answer to a why question.
- Statement connects events and actions in the story that are consecutive or more distant (e.g., sentences at the beginning and end of the story) by explicitly stating or strongly suggesting causal relationships between them

- Statement expresses an internal state (desires, feelings, thoughts, attitudes) that serves as either cause or effect of another state, event, or action
- Statement may involve using world knowledge to determine cause or effect or linking two parts of text by relating them in a cause-effect relationship
- Statement provides an answer to 'what will/might happen next?'
- Statement provides information about what character(s) intend to/plan to do
- Statement explicitly refers to future events (with a consequential quality)
    - May be expressed in certain terms (e.g., going to; gonna; will)
    - May be expressed in less certain terms (e.g., I guess; might; should; could; probably will; maybe will/going to, etc.)
- Statement scored regardless of whether it is substantiated by future text or not

Informational

- Statements provide answers to what, how, who, where, when questions
- Statement involves an inference about states, events, objects, or actions expressed in the focal sentence but does not involve cause and effect relationships
- Such statements include
    - Elaboration of event or actions
    - Elaboration of words or concepts
    - expressions of manner (usually indicated by adverbs);
    - descriptions of objects (qualitative or quantitative but not truly evaluative)
    - descriptions of characters (including actual and potential internal states)
    - immediate locations (locatives such as behind, under);
    - identification and description of settings (place, location, destination);
    - specification of spatial information;
    - specification of temporal information (time, season, etc.)

Evaluation

- Reader makes a judgment/evaluation of a state, event, action, or object described by the focal sentence based on world knowledge, social conventions, or own values.
- e.g., (from JAO_TA2 Gabby [9.1]) "it's good that they were careful"
    - 2 points: statement in which the reader's evaluation or judgment is explicit (e.g., "That's good/that's bad/oh no! That's dangerous!")
    - 1 point: statement in which the reader's evaluation/judgment is implicit (e.g., "Oh no!/wow!")

Uninterpretable

- Coder cannot interpret statement due to semantic, phonological, and/or syntactic errors
- Do not assign points for uninterpretable statements (count only as total number of uninterpretable statements)