# UPDATING SKILLS AND CHARACTER REFERENCING IN CHILDREN WITH SPECIFIC LANGUAGE IMPAIRMENT

by

JILLIAN MELISSA FRICK

B.A., The University of Saskatchewan, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2012

## Abstract

Specific language impairment (SLI) involves unexpected delays in language development that are deemed primary, rather than secondary to other developmental issues. This study investigated executive functioning and referencing abilities in children with SLI. Referencing is a complex task that may particularly depend on working memory (WM) resources, specifically in terms of updating abilities. Updating entails the active manipulation of WM contents by replacing older information with newer and more relevant information. If both linguistic abilities and processing resources are critical for adequate referencing, then children with SLI would be particularly vulnerable in this area. There is limited evidence regarding updating skills in children, referencing abilities in SLI, and the possible influence of updating on language production.

A group of 12 children with SLI aged 5 to 8 years were matched to 12 same-age typically developing (TD) peers. The children completed updating, short-term memory (STM), and story-telling tasks. Compared to the children with TD, the children with SLI performed significantly poorer on all updating tasks and also consistently exhibited lower overall referential adequacy, but whether this applied to both nominal and pronominal forms depended on the story. There were also many parallels, however, between the groups in terms of referencing. Specifically, each group responded similarly to story differences and to the demands of the various referential functions (introduction, maintenance, and reintroduction) in terms of the frequency of character references, the referential types selected, and patterns of referential adequacy. Regarding the relationship between WM and referencing, visual STM was the only memory task that correlated significantly with referential adequacy in the SLI group. In contrast, updating did correlate significantly with referential adequacy in the TD group. It is still possible that updating plays a role in the ability to adequately refer to characters in both groups, but this relationship could be mediated or overshadowed by limitations in basic storage capacity or by linguistic factors for

children with SLI in this developmental window. Due to the small number of participants in this study, the results regarding the relationship between updating abilities and referential adequacy are tentative and require replication with a larger sample.

## Preface

The following research was approved by the UBC Behavioural Research Ethics Board. The Certificate Number of the Ethics Certificates obtained was H11-00806.

# Table of Contents

# List of Tables

# List of Figures

## Acknowledgements

First and foremost, I must thank Paola Colozzo for her incredible guidance and devotion as my supervisor. I cannot begin to measure the hours and effort you invested in this project, and I admire you greatly as a researcher and teacher. Cristy Whitely also played a significant role in designing this project, and I thank her for patiently passing her knowledge and ideas on to me. Thank you to Valter Ciocca and Jeff Small for providing valuable direction as my committee members. I deeply appreciate both Cristy Whitely and Heather Morris for teaching me the details of data collection, analysis, transcription, and coding, and for performing plenty of these duties themselves as well.

The speech-language pathologists who helped to recruit participants were critical to this project. The parents and teachers who agreed to, and arranged for, the participation of their children and students were also very helpful. And to the children who enthusiastically participated in this study – everything, of course, was only possible thanks to you! Finally, I am beyond grateful to my family, friends, and boyfriend for their endless support.

# 1 Introduction

This study contributes to our understanding of both linguistic and information processing abilities in children with specific language impairment (SLI). It considers updating of working memory (WM) and a particular aspect of narrative production in children with SLI and age-matched typically developing (TD) peers in kindergarten to grade 2. Narrative production is a complex task, requiring the speaker to construct and keep track of the plot's details and to monitor the listener's needs. One aspect of cohesion in a narrative is referencing, the use of linguistic devices to refer to characters. Adequate referencing to create an unambiguous message draws on many different abilities in one's linguistic and cognitive systems. Referencing may particularly depend on WM resources, specifically in terms of updating abilities.

The following introductory chapter will first describe the complex task of narrative production. Second, a specific type of cohesion, personal referencing, will be defined in terms of requirements for adequate use. Third, the potential contributions of WM and updating to referencing skills will be described. Fourth, the criteria for SLI will be defined, and existing research on this population's referencing and WM abilities will be presented. Finally, the current study's research questions will be presented along with their corresponding hypotheses.

## 1.1 Narrative text

Text is a basic unit of spoken or written language meaningfully organized as a unified whole (Beliavsky 2003; Liles, 1985). Narrative discourse is one dynamic genre of text that represents events, real or imagined, in related utterances (Bennett-Kastor, 1983). Telling stories is natural and pervasive in the daily lives of young children – in social interactions (i.e., personal narratives), as recreational activities (e.g., books and television), and in the classroom (Schneider, Hayward, & Dubé, 2006). Narrative production is a complex task. It requires the speaker to continuously construct the plot's details, consider the context, keep track of what he/she has said, and monitor the listener's current understanding and attentional state.

To fulfill all of these requirements, storytelling draws on many different linguistic, cognitive, social, and pragmatic abilities. Linguistic skills are needed to capture semantic relationships through adequate grammar and vocabulary. The successive chain of events must be tied together in a structured and meaningful unit, while applying grammatical structure to the content as it is presented (Wigglesworth, 1990). Cognitive skills include memory, symbolic representation, sequencing, and self-monitoring to keep track of referents and to create a coherent and cohesive text. The storyteller is required to monitor, integrate, store, and retrieve large amounts of information (Montgomery, Magimairaj, & Finney, 2010). Social and pragmatic skills include topic maintenance and assuming another person's point of view. The narrator must differentiate given information from new information, and understand the listener's current knowledge state and needs. All of these critical aspects occur in real time, simultaneously and continually throughout the narrative.

Telling a narrative is thus a demanding task that captures valuable information about many facets of a child's communication abilities (Liles, 1993). For assessment, researchers and clinicians are moving beyond words and sentences to the longer units of the narrative (Schneider et al., 2006). Many researchers argue that narrative production is a more valid indicator of language development and use, especially since it is a more natural, dynamic, and spontaneous task that taxes verbal and nonverbal abilities alike (Heilmann, Miller, & Nockerts, 2010; Liles, 1993). Compared to conversations, narratives tend to incorporate longer utterances and more complex language (Heilmann et al., 2010). The child also does not get to share the floor to take a break, as would be an option in conversation (Colozzo, Gillam, Wood, Schnell, & Johnston, 2011). Narrative abilities have also proven to be a particularly sensitive predictor of later language, literacy, and academic outcomes (e.g., Botting, Faragher, Simkin, Knox, & Conti-

Ramsden, 2001; Miller et al., 2006; Montgomery et al., 2010; Roth, Spekman, & Fye, 1995; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998; Tomblin, Records, & Zhang, 1996). Researchers have considered many aspects of narrative. The current study focuses on character referencing in narratives.

## 1.2 Personal referencing

Warden (1976) likened referring expressions to verbally pointing to an item. *Referencing*, the use of linguistic devices to refer to an item within or outside of the text, is one specific category of cohesion (Beliavksy, 2003). Cohesion is an aspect of narrative organization that involves using linguistic devices to establish connections within and between utterances (Halliday & Hasan, 1976). Cohesive ties occur when one element in a text depends on another element to be interpreted (Baltaxe & D'Angiola, 1992; Finestack, Fey, & Catts, 2006; Roth et al., 1995). These ties link the text together as a meaningful, unified whole, supporting the narrative's overall flow and organization (Bennett-Kastor, 1983). Referencing creates a chain of continuity between characters, places, things, and events that are mentioned in a narrative more than once (Baltaxe & D'Angiola, 1992). The distance and direction between a cohesive tie and its referent can vary – the link can refer to an item mentioned either previously (anaphoric reference), subsequently (cataphoric reference), or even outside of the text in cases of exophoric referencing (Hickman, 1980; Karmiloff-Smith, 1985). Linguistic devices such as personal pronouns (e.g., *he*), possessive pronouns (e.g., *his*), determiners (e.g., *a* girl*, the* dog), demonstratives (e.g., *that* boy), and comparatives (e.g., the *taller* boy) are used to make reference in discourse (Roth et al., 1995). Personal referencing indicates a referent's identity, specifying features such as gender, number, and animacy (Halliday & Hasan, 1976). There is a great amount of information, therefore, that children learning language must represent in the referential devices they choose.

The concept of a discourse model has been proposed by Johnson-Laird and Garnham (1980). It corresponds to a mental representation of relevant characters, events, relationships, discourse rules, intentions, and knowledge. A model is created and modified throughout the discourse production. The speaker and the listener each construct and modify their own separate discourse models based on their knowledge states and beliefs, although these models interact. The speaker considers the listener's discourse model when constructing a narrative, while the listener interprets what the speaker has said. This is an active cycle in which the discourse models influence the interpretations of a narrative, while these interpretations also continually modify the models (Isard, 1974; Levelt, 1989). It is assumed that the listener will interpret a narrative more successfully when there are fewer discrepancies between the discourse models.

Effective referencing requires a speaker to take the listener's discourse model into account. The speaker must distinguish characters and organize events with the listener's perspective in mind (Liles, 1985). To decide what information the listener needs, the speaker must monitor the listener's current knowledge, expectations, and attentional state regarding each referent, as well as the function of the reference (Karmiloff-Smith, 1985; Maratsos, 1974; Roth et al., 1995; Wong & Johnston, 2004). Distinguishing between information that is given (i.e., familiar and assumed) versus new, thus requiring more specific explanation, can be pivotal to successful discourse (Warden, 1976). The referential forms that are employed will be based on making this distinction and on successful monitoring of both the listener's state and the references produced up until that point (Bennett-Kastor, 1983).

Gundel, Hedberg, and Zacharski (1993) discussed a concept similar to a discourse model to account for the way referential forms are used by the speaker and interpreted by the listener. Each referential token takes on a cognitive status that is dependent on the speaker's assumption of the listener's knowledge and attentional state in regards to the referent. The mental representation of a listener's knowledge can include general cultural knowledge and whether the referent has already been introduced. Different referential forms encode and signal different assumed cognitive statuses of the referent. A givenness hierarchy from most to least restrictive referential forms (low to high on the givenness hierarchy) determines whether a cognitive status has been met if a certain form is used. For example, *familiar* is a cognitive status high on the givenness hierarchy in which the intended referent is represented in the listener's memory and thus can be uniquely identified. This is necessary in order for a personal pronoun (e.g., *him*) or a definite demonstrative (e.g., *that*) to be used. In contrast, *type identifiable* is a cognitive status lower on the givenness hierarchy in which the listener is only able to access a representation of the type of thing being referred to. This requires an indefinite expression (e.g., *a* dog) to be used. According to Gundel and colleagues, cognitive statuses and the givenness hierarchy interact with Grice's (1975) Maxim of Quantity when the speaker selects an appropriate form: the maxim states that a speaker should make contributions as informative as necessary while not providing more information than required. In order to choose the appropriate forms for referencing to create a complete reference, the narrator must monitor the listener's knowledge and attentional state in regards to the referents as the story unfolds.

A complete reference refers to an entity whose identity is clearly specified in the text. In the case of personal referencing in a fictional narrative, the speaker must learn the rules of referencing to specify and differentiate the various characters by appropriately mapping the linguistic form to the current function (Bamberg, 1987). Character referencing can serve one of three different functions. First, *introduction* involves the first mention of a previously unidentified character. For introducing a character, a proper noun or indefinite article (e.g., *a* boy) should generally be used. Use of nonspecific referencing through an indefinite article implies no particular member of a class, as the character is being identified for the first time (Maratsos, 1974; Warden, 1976). Second, *maintenance* involves continuing reference to the same character in subsequent utterances. For maintaining reference to a character, a definite article (e.g., *the* boy), a pronoun (e.g., *he*, *she*), or a zero form (i.e., subject ellipsis; e.g., "*The boy* cut the wood and then Ø painted it.") are appropriate. A definite article indicates that a particular character is already uniquely established and activated in the narrative (Emslie & Stevenson, 1980). Pronouns and ellipses have little descriptive content, so the referent needs to already be activated in order to be identified by the listener (Gundel et al., 1993). Third, *reintroduction* brings back a character that was temporarily out of focus over one or more utterances. For reintroducing a character, a proper noun or a definite article should generally be used.

The requirements of updating the discourse model, monitoring the needs of the listener, and keeping track of characters and functions suggest that certain cognitive resources will be taxed when performing these tasks.

## 1.3 Working memory and updating

As we have established above, skills in referencing are related to linguistic, cognitive, social, and pragmatic abilities (Baltaxe & D'Angiola, 1996). All of these abilities combine to provide an unambiguous message through the complex organization of ideas, which is a process that can be taxing for our limited capacity system (Baltaxe & D'Angiola, 1992; Hoffman & Gillam, 2004). As will become clear below, referencing may be particularly demanding of WM resources, specifically in terms of updating abilities. Updating involves dynamically monitoring incoming information and coding for its current significance (Miyake et al., 2000). It entails the active manipulation of WM contents by replacing older information with newer and more relevant information, which rests on continually adjusting activation levels (Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). Updating is one of the processes postulated to contribute to executive functioning.

Executive functioning is a general control mechanism presumed to play a key role in complex cognitive processes (Lehto, 1996; Miyake et al., 2000; van der Sluis, de Jong, and van der Leij, 2007). Although there is no clear consensus regarding how or whether executive functioning is fractionated and whether it is a domain-general resource, three executive functions (namely updating, shifting, and inhibition) have received the most attention in research to date (Friedman et al., 2008; Lehto, 1996; Miyake et al., 2000; Shah & Miyake, 1996). These executive functions have been found to be separable and distinct, yet moderately correlated. This suggests both diversity and unity, perhaps involving some common mechanisms since the functions are not completely independent (Lehto, 1996; Miyake et al., 2000).

In the WM model proposed by Baddeley and Hitch (e.g., Baddeley & Hitch, 1974; Baddeley, 2003), executive functioning would correspond to the roles of the central executive (CE). In this model, the CE is a domain-general (i.e., modality-free), limited-capacity supervisory mechanism that flexibly controls and regulates information flow. The WM system includes two domain-specific specialized 'slave' subsystems: the phonological loop and the visuospatial sketchpad. The phonological loop rehearses and stores sound-based material, including verbal material. The visuospatial sketchpad processes and retains visual and spatial information and mental images (Baddeley & Logie, 1999). The WM system is responsible for continuous and simultaneous monitoring, processing, manipulating, and maintaining relevant information stored temporarily in memory (Baddeley & Logie, 1999). To accomplish this, limited resources are flexibly distributed (Gathercole, 1999; Montgomery et al., 2010). WM capacity changes with development and varies between individuals (Gathercole, 1998, 1999).

Recent studies have begun to show that CE capacity, including updating, improves with age (Mantyla, Carelli, & Forman, 2007; Whitely & Colozzo, under review). Gathercole (1998) found significant age-related capacity increases in both short-term memory (STM) and WM tasks, the latter presumed to involve the CE. Children's memory function improved between infancy and the early school years. It was close to adult abilities by 7 years of age, but continued to improve into adolescence (Gathercole, 1998). Regarding updating in particular, Mantyla and colleagues (2007) gave 51 children (aged 8 to 12 years) and 62 young adults time-based prospective memory tasks, in which they performed updating tasks while monitoring a deadline. The younger children had lower performance in the updating task compared to the young adults.

Recent studies have also suggested that updating is domain general. For example, St Clair-Thompson and Gathercole (2006) used exploratory factor analysis to study executive functioning in children 11 to 12 years of age ($N = 51$). The authors concluded that updating abilities were associated with both verbal and visuospatial complex WM tasks. Van der Sluis and colleagues (2007) also found that updating was positively related to both verbal and nonverbal reasoning, reading, and arithmetic. The authors used confirmatory factor analysis to explore the variance of multiple executive function tasks in a large sample ($N = 172$) of 9- to 12-years-olds.

Executive functioning, including updating, is deemed important for learning complex skills in childhood (St Clair-Thompson & Gathercole, 2006). Attention and processing resources must be actively managed to select, formulate, keep track of, and integrate relevant content (Carretti, Cornoldi, De Beni, & Romano, 2005; Im-Bolter, Johnson, & Pascual-Leone, 2006; Karmiloff-Smith, 1985; Palladino et al., 2001). Information can be lost due to decay or interference if cognitive processing is slow and if representations are weak (Montgomery et al., 2010). This poorer processing can build a disorganized and impoverished language network, which can in turn make the updating of relevant information less efficient (Im-Bolter et al., 2006; Mantyla et al., 2007). Reading skills are also affected by the strength of memory representations, the speed of information retrieval, and the ability to attend to what is relevant (van der Sluis et al., 2007). Reading comprehension requires one to continually hold and update information about the text in WM (Palladino et al., 2001).

Palladino and colleagues (2001) explored the role of updating in reading comprehension in children and adolescents (aged 11 to 15 years, $n = 60$), as well as adults (aged 19 to 21 years, $n = 127$) in a series of five experiments. The participants completed STM, WM, updating, and reading tasks. The authors found that participants with poorer reading comprehension had poorer verbal WM, updating skills, and recall than those with better reading comprehension, whereas the groups did not differ in STM skills. The participants with poorer reading comprehension also made more intrusion errors, in which they remembered more non-target items, compared to the participants with good reading comprehension. Error rate was also positively correlated with increased task demands, both in terms of the size of the pool (i.e., total number of possibly relevant items) and the number of items to remember. Carretti and colleagues (2005) also found that updating and inhibition abilities mediated the relationship between WM performance and reading comprehension in a large sample ($N = 218$) of children aged 8 to 11 years. The children with poorer reading comprehension consistently had more inaccurate recall, intrusion errors, and omissions on the memory tasks. A second experiment confirmed that these results were the same one year later, and that the findings generalized to other updating tasks. The authors cautioned, however, that their findings do not indicate that the relationship between updating and reading comprehension is causal.

There is limited evidence regarding language production and updating. This goal-directed, flexible manipulation of WM would nonetheless seem particularly relevant for referencing. As discussed above, clear referencing would require one to constantly update the discourse model and to keep track of characters in order to select the right form according to the function. A recent study by Whitely and Colozzo (under review)[1] specifically explored updating skills and referencing in narratives. They found that updating and referential adequacy were

[1] This study by Whitely (née McNiven) and Colozzo built on the thesis of McNiven (2010). The sample size was increased and all analyses were redone.

significantly related in a sample of 64 children with typical development (TD) from kindergarten to grade 2. This relationship continued to be significant when the effects of age and of basic STM capacity were controlled for statistically. The results suggest a relationship between updating and referencing abilities in TD children in this developmental period.

Although all functions should depend on updating resources, switching, which includes the functions of introduction and reintroduction, is presumably more demanding than maintenance (see section 1.3 above for descriptions and examples of the various functions). This is because it requires the speaker to choose the appropriate linguistic device based on his/her assessment of the listener's knowledge of the character. For introduction in particular, this function can challenge a child's undeveloped differentiation of indefinite and definite expressions (see section 1.5, below). For reintroduction, the distance between the referent and its identifier is increased (Daneman & Carpenter, 1980). The constant monitoring and revising of relevant information associated with switching likely taxes one's updating abilities. Maintenance, in contrast, would seem to demand less of one's updating resources because the referent stays in focus over successive utterances. These predictions were partially borne out in Whitely & Colozzo (under review), as both switching functions proved to be more difficult than maintenance for the sample of 5- to 8-year-olds. Only the relationship between maintenance and updating, however, proved to be significant for these children. The authors hypothesized that maintenance was most sensitive to differences in updating abilities because even this comparatively easier referential function was exacting for the children. The length of the stories and the multiple characters that the children had to keep track of likely increased the demands of maintaining clear reference to characters for these young storytellers. The switching functions (introduction and reintroduction) might have been so demanding for the children that they recruited non-executive resources (i.e., reflected by the STM tasks) to an extent that overshadowed the role of updating.

Thus, the few existing studies on the relationship between updating and linguistic and reading skills suggest that updating increases with age, may be domain general, and is correlated with higher-level cognitive tasks, including those that involve language. Updating in particular may be critical for personal referencing. The present study aimed to add to the sparse literature on updating and language *production* in children by focusing on those who are likely to show deficits in both of these areas – children with language impairments. The next section considers the literature regarding the development of referential abilities in typical children, before we move on to review the research regarding children with language impairments.

## 1.4 Development of referential abilities in typically developing children

Narrative production by TD children has been an interest in research for the last 30 years (Liles, 1993). Cohesion in particular has been found to become more adequate as children get older, and may become quite stable by about age 7.5 years, although this varies considerably depending on the demands of the specific task (Liles, 1985). Referencing skills also improve with age, in terms of both the quantity and the adequacy of the forms produced (Beliavsky, 2003; Wong & Johnston, 2004). With increasing age, children become more successful at clearly referring to characters, time, places, and things in their stories. This development may be at least partially attributable to changes in updating capacity (Whitely & Colozzo, under review). Referencing abilities also depend on social language skills, and on developing and monitoring a discourse model (Baltaxe & D'Angiola, 1996). Finally, referencing also relies on linguistic knowledge, with particular emphasis on mastering the use of determiners and of pronouns for the specific purpose of referencing.

A consensus has yet to be reached regarding specific ages at which referential devices are acquired, although the existence of a regular developmental trend is agreed upon (Strong & Shaver, 1991; Wigglesworth, 1997; Wong & Johnston, 2004). Young children initially depend on the context and on extralinguistic gestures (e.g., point, eye gaze) for referencing (Karmiloff-Smith, 1980). Before 5 years of age, children introduce new referents using forms that assume the referents to be already known to listeners (Wong & Johnston, 2004). By kindergarten, most children are aware of pronoun rules and can judge the listener's knowledge, yet errors still occur (Wong & Johnston, 2004). At this stage, children cannot consistently manage referential forms in narratives with more than one character or of increasing length (Bamberg, 1987; Karmiloff-Smith, 1980). Bennett-Kastor (1983) gathered stories from 12 children between the ages of 2 and 5 years. The authors found that the 5-year-olds told longer stories involving greater cohesion and complexity of noun phrase use compared to children only one year younger. Considering referencing in older children, Hickman (1980) found that 10-year-olds consistently used appropriate referential forms more often than did 7-year-olds when narrating films for an unfamiliar listener. By the ages of 7 to 10 years, children are apparently starting to be quite flexible in their use of referential devices. The developmental trends found in the studies mentioned above were confirmed by the work of Wigglesworth (1990), who studied how children ages 4, 6, and 8 years referred to characters for different functions, compared to adults. The results showed that no child performed at the adult level for referential control. Referencing by the 4-year-old children tended to be ambiguous, using deictic forms of reference (i.e., pointing), although attempts to create cohesive narratives were evident. The 6-year-old children showed more conscious understanding of and concern for referring expressions, and overall organization of their narratives. They made more self-repairs (i.e., backtracking to make corrections) to their references than did any other age group, suggesting conscious awareness of referencing skills that were still being acquired. The 8-year-old children produced more complex and detailed narratives, but did not perform at the adult level.

In another study, Wigglesworth (1997) found that children tended to progress through numerous stages of referential organization. Three general referential strategies used by children were analyzed: thematic subject (i.e., preferring pronominals for all functions to refer to the main protagonist), nominal (i.e., preferring nominals for all functions to refer to all characters), and anaphoric (i.e., preferring nominals for switching and pronominals for maintaining reference). Control over strategy use increased with age, although the stages were not linked to specific ages. Varying the complexity of the narrative segments affected the cognitive load of referencing, which led to the use of various strategies by all age groups. The referential load of each segment was based on the number and activity of the characters available at a given time to be included in the story. Increased referential complexity led younger children to adopt simpler strategies, while older children were better able to maintain the same strategy when managing different loads. The 4-year-old children lacked consistency in strategy use, but created a coherent plotline. They also used deictic gestures for reference, which meant that the listener would require the pictures for understanding. The 6-year-old children simplified their stories when strategies were maintained in more challenging conditions. The 8-year-old children varied the most in strategy use, and mostly used specific and nonspecific referents correctly. The 10-year-old children achieved overall narrative organization and maintained their strategies in varying conditions, indicating that they were less affected by the referential load compared to the younger children. Their stories, however, were still less complex than those told by the adult group, as the adults were able to maintain clear and unambiguous referencing.

Effects of task demands on referential adequacy were also found by Schneider and Dubé (1997) in 44 children in kindergarten and grade 2. They compared referential adequacy for introduction and maintenance in story retells based on three story presentation conditions: oral-only, pictures-only, and oral-with-pictures. Oral presentation provides a storyline and linguistic structuring, but challenges WM. Visual presentation, on the other hand, supports memory but requires the translation of visual information into verbal information and the independent formulation of a storyline without a linguistically structured model. The results showed a significant effect of condition, with referential adequacy being highest for both grades of children in the oral-only condition, and lowest in the pictures-only condition. There was also a significant effect of age, as the older children had higher referential adequacy than the younger children in all conditions. Finally, there was a significant interaction between condition and age, as the younger children had higher referential adequacy in the oral conditions whereas the older children's referential adequacy was equally high across conditions.

Mastering the indefinite/definite distinction impacts the adequacy of a reference as well. Through a series of three experiments in various discourse contexts, Warden (1976) concluded that children younger than age 5 had significantly more inappropriate uses of the definite article in English. The youngest children, the 2-year-old group, frequently used deictic reference. Accurate use increased but remained inconsistent between 5 to 9 years of age. Nearly appropriate contrastive use of definite and indefinite reference was evident by age 9, but performance by these children was still less accurate than by the adult participants. Overall, use of the indefinite article was mastered earlier than the definite article, and article use was affected by the task context. Roth and colleagues (1995) also found that the use of definite versus indefinite articles for referencing became quite accurate by 8 to 10 years of age, but that development of pronominal use continued up to age 14.

Some researchers have suggested (e.g., Warden, 1976) that egocentricity (i.e., not having the ability to consider the listener's point of view) could explain the difficulties with referential adequacy faced by younger children. Emslie and Stevenson (1980) found, however, that non-egocentric article use was mastered by 3 years of age, and thus proposed that inadequate referencing may simply be due to undeveloped discourse skills or high cognitive demands in the narrative task. Maratsos (1974) also concluded that competence in distinguishing when to use specific or nonspecific articles is established early for children. He told the beginning of a story and then asked children (aged 3 and 4 years) questions requiring a choice of reference. Maratsos found a reliable age effect in which the 3-year-old children used definite noun phrases erroneously when elicited by the questions. Referencing by the 4-year-old group was more accurate, although the group was split into low and high performance groups. Overall, indefinite reference was established early, while definite reference improved from ages 3 to 4. It should be noted, however, that the task only involved isolated referencing in response to adult questions.

Overall, existing research on the narratives of TD children shows that referential adequacy increases with age, and it also varies by referential function. Children initially tend to rely on extralinguistic cues to specify a character. Early use of linguistic devices contains errors likely due to incomplete mastery of linguistic knowledge and the demands of creating and updating a discourse model. Research suggests that errors are not likely due to lack of consideration of the listener's perspective. Specific ages of mastery vary across studies, likely due to task differences. Yet so far it is apparent that children are still progressing in their referential skills into the early school years. The definite/indefinite distinction and supplying sufficient information across utterances are particularly challenging. Children seem to have better control earlier of the indefinite form, but this could simply reflect the default choice in young children. Therefore, it is suspected that young children would struggle the most with the referential functions of introduction and reintroduction as these functions are more likely to require the differential use of these forms.

One aspect of referential adequacy that has received less attention is how improvements with development vary depending on the various referential functions. The limited studies that have contrasted the various referential functions have found two important trends. First, the types of referential forms used vary by referential function (Wong & Johnston, 2004) with, on the one hand, greater pronoun use for maintenance and, on the other hand, greater use of referential forms lower on the givenness hierarchy (e.g., noun phrases) for switching (i.e., introduction and reintroduction). Second, children have more difficulty with switching compared to maintenance. As discussed above, this is likely attributable to both linguistic factors (e.g., indefinite and definite determiners, pronouns) and demands of discourse model and updating resources (e.g., Bamberg, 1987; Whitely & Colozzo, under review; Wong & Johnston, 2004). Next we will consider the relevant research on these topics with regards to children with language impairments. If both linguistic abilities and processing resources are critical for adequate referencing, then children with SLI would be particularly vulnerable in this area. The following sections will provide general information regarding children with SLI, and will review the existing literature pertaining to referencing and WM in this population.

## 1.5 Specific language impairment

Children with SLI have unexpected delays in language production and/or comprehension that are deemed primary, rather than secondary to other developmental issues. The involved language deficits are heterogeneous within this group, and do not have a known cause. These children have intact general physical, nonverbal cognitive, social-emotional, and hearing functioning, as well as typical exposure to language (Archibald & Joanisse, 2009). There is substantial evidence that children with SLI are at risk for lifelong negative educational, social, and emotional difficulties (Tomblin et al., 1997). The need for consensus on valid, reliable, and objective diagnostic criteria for identifying SLI is recently receiving increased interest (Heilmann et al., 2010; Tomblin et al., 1996). Children with SLI form a heterogeneous group, so multiple factors are likely to be at play (Ellis Weismer, Evans, and Hesketh, 1999). Researchers and clinicians must remember that SLI occurs on a continuum, so caution should be taken in categorizing children. Narratives have been the focus of much research on children with SLI, but referencing has been studied to a lesser extent.

### 1.5.1 Referencing abilities in children with SLI

Children with SLI often struggle with various aspects of narrative production, from the global text level to the sentence level, and in terms of content, form, and organization (e.g., Colozzo et al., 2011; Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Liles, 1993). Some children's story production scores can help to predict their risk for SLI, as well as later oral language development and even future academic achievement (Schneider et al., 2006). Many studies that have considered either content or structure in the narratives of children with SLI have found that both aspects can be problematic (see Colozzo et al., 2011, for a review). For example, Colozzo and colleagues (2011) analyzed both the content and form of fictional stories generated from picture stimuli by 76 school-aged monolingual children. The narratives by the SLI group had less adequate content (e.g., fewer utterances, fewer story elements) and form (e.g., more grammatical errors, less syntactic complexity) compared to age-matched peers. Dissociations between content and form were evident in the majority of the SLI group's narratives as well. Most children with SLI fell into one of two groups, with either relatively better content and low grammaticality or unelaborated content and fewer grammatical errors. They concluded that children with SLI might have difficulties with the combined demands to create a story that is both elaborate and grammatically accurate.

A few studies have focused more specifically on cohesion, including referencing, by children with SLI. Existing data on the referential abilities of children with SLI is limited, but has consistently indicated that children with SLI used less adequate cohesive ties than did same-age peers (Finestack et al., 2006; Liles, 1985; Norbury & Bishop, 2003). In fact, some researchers have suggested that referential abilities may be one of the most discriminating measures to differentiate children with SLI from their TD peers, and that this could be an important area to consider for intervention (Fey et al., 2004; Liles, 1985). One example of poorer referential abilities in children with SLI is a study by Liles (1985), who compared the narratives of 20 children with TD and 20 children with mild to moderate language disorders (LDs), aged 7 to 10 years. The narratives were told to adults who viewed a movie with the child (shared condition) and adults who had not viewed the movie (unfamiliar condition). The LD group produced narratives that included fewer cohesive elements and personal references. They also produced more erroneous and incomplete cohesive ties and more exophoric referencing. It is interesting to note that both groups similarly modified their use of cohesion according to the listener condition. When the listener had not already seen the movie, both groups provided a significantly higher percentage of complete ties. This similarity suggests that the children with LD perceived the listener's need for information, so that deficient referencing was not due to egocentricity. The results of a study by Schneider and Hayward (2010) suggested that TD and SLI group differences in referential adequacy of first mentions persist into the early school years. The authors investigated referential adequacy for the function of introduction in the narratives of children aged 4 to 9 years: 300 children with TD and 77 children with SLI. A significant age effect showed a gradual increase in referential adequacy from ages 4 to 7 in both groups, with little change between ages 7 to 9. Significant group differences were also found, with the SLI groups obtaining lower scores compared to the TD groups for all ages except the 9-year–olds.

Although referential adequacy has been shown to be lower for children with SLI, other characteristics of their referencing are more similar to children with TD, including rates of usage, error patterns, and type by function distributions. In terms of rates and error patterns, Baltaxe and D'Angiola (1996) analyzed referencing in 1-hour play samples from children with SLI (mean age 7;7) that were matched for receptive and expressive language age (determined by mean utterance length, receptive vocabulary, and syntax) to younger TD children (mean age 3;5). The groups did not significantly differ regarding the rates and patterns of usage of referential forms nor in terms of error rates. Although referencing was less successful for the SLI group, the differences between the groups were nonsignificant. The SLI group's errors were primarily omission errors and identification errors. In terms of distribution of type according to function, Norbury and Bishop (2003) looked at referencing for the two main characters in the narratives produced from a wordless picture book by 6- to 10-year-old children with SLI and TD. Plural referents were excluded from all analyses. Both groups were similarly affected by function in their referential type use, and tended to use nominals (i.e., proper nouns or noun phrases) more than pronouns overall. For introduction, the TD group correctly used an indefinite noun phrase 67% of the time, compared to the SLI group who did so 53% of the time, and none of the children ever used a pronoun. The children in both groups also very rarely used pronouns for reintroduction, as approximately 95% of references for reintroduction consisted of nominal expressions. In contrast, both groups preferred to use pronouns for maintenance; nominals were used for 17% and 29% of cases where reference to a character was maintained by the TD and SLI groups, respectively. Ambiguous pronouns were not included in any of these totals, however. Considering ambiguous pronouns for reintroduction and maintenance combined, there was a statistically significant difference between the groups. Ambiguous referencing using pronouns occurred in 12% of all referencing by the TD group, compared to 41% by the SLI group. The group with SLI had more difficulty with maintenance and reintroduction, as indicated by the higher levels of ambiguous pronoun use. Finestack and colleagues (2006) also found that children with SLI did not differ from TD children for referential rate, although their referential adequacy was poorer. They gathered narratives from 569 children in grades 2 and 4, with a wide range of verbal and nonverbal abilities. Narrative production using a picture set was used to study the frequency and completeness of pronominal referencing. First, an examiner provided modeling by labelling key elements of a story for one picture set, and by reading a story corresponding to another picture set. Then each child generated one oral narrative and one written narrative based on two more sets of coloured pictures. The authors found that the SLI group had a significantly lower rate of complete pronominal referencing compared to the TD group. The SLI group did not, however, significantly differ from children in the TD group in terms of their rate of pronominal referencing. The most critical issue with this study was the removal of the children who did not produce five or more pronouns in their narrative sample from the group comparison analyses. This resulted in the removal of 372 (65.4%) participants from the sample. This likely overestimated the referencing skills of the group overall as it may have created a ceiling effect. The authors hypothesized that the narrative generation task might have been too difficult for all of the children due to the cognitive processing required to generate a story. The overall length of the testing sessions (2 hours) may have been too long and taxing for the children as well. Nonetheless, this study suggests differences between the groups, not in referential rate but in adequacy.

In general, research on referencing by children with SLI suggests that they do not differ from TD children in terms of rates of referencing, distributions of referential types, and types of errors. The difference between groups is that children with SLI have lower referential adequacy overall. Research findings do not support the explanation that this is due to failure to take the listener's perspective. The generalizability of the research findings in this area is limited, however, due to variability in SLI criteria, small sample sizes, the specific cohesive ties studied, different age groups and matching criteria for the participants, and/or differences in the narrative tasks.

### 1.5.2 Memory abilities in children with SLI

Children with SLI may have difficulties with referencing due to a combination of factors. Constraining factors could include linguistic knowledge and abilities, as well as more general factors such as processing speed and WM capacity. Language and cognition interact during development, and there is much evidence that children with SLI have more limited WM resources (see Montgomery et al., 2010, for a review). More research is required to understand the role of cognitive processing and its specific mechanisms in language production and comprehension (Charest & Johnston, 2011; Charest, 2012; Colozzo et al., 2011). Gillam and Hoffman (2004) have described the relationship between language and cognitive development as a loop in which information processing is required for language, and language in turn further supports complex information processing – hence cognitive abilities both underlie and are constrained by language issues (Gillam & Hoffman, 2004).

In the last twenty years, numerous studies have explored the links between WM, information processing, and language limitations in SLI, yet many questions remain. A literature review by Montgomery and colleagues (2010) highlighted that many children with SLI have shown limitations in WM, verbal STM, processing speed, and CE functions (e.g., updating and inhibition). Visual STM tended to be more similar to TD children. Still, Montgomery and colleagues suggested that many children with SLI show difficulties across all mental processes and modalities. Not all of the children with SLI studied in the literature review exhibited WM deficits, however, and more research on the specific association between WM and language deficits is still needed. Finally, few studies have focused specifically on updating of WM, although studies have explored CE more broadly by considering performance on other complex WM tasks (e.g., listening spans, dual tasks).

There is evidence that children with SLI often have poorer functioning verbal WM than age-matched peers with typical language, and even language-matched younger children (e.g., Alloway, Rajendran, & Archibald, 2009; Briscoe & Rankin, 2009; Gathercole, 1999). For example, Archibald and Gathercole (2006) had 20 children with SLI aged 7 to 11 years complete a test battery consisting of various WM and STM tasks. They found a "double memory jeopardy" in which children with SLI showed significant deficits in both complex verbal WM (95% had deficits) and verbal STM (70% had deficits) tasks compared to same-age peers. Half of the SLI group also showed differences compared to peers in visuospatial STM tasks, although their scores were still in the average range. There were no nonverbal complex WM tasks included in the study.

Alloway and colleagues (2009) also found that children with SLI in their study had deficits in verbal WM. Fifteen children with an average age of 9.2 years performed a variety of memory tasks. Only a few children with SLI scored below age norms on complex visuospatial WM and visuospatial STM tasks. As noted in the Archibald and Gathercole (2006) study above, findings tend to be mixed in regards to visuospatial memory skills. Archibald and Joanisse (2009) once again found deficits in a verbal STM task, consistent with many other studies. But only some of the children with SLI had verbal WM impairments, which suggests that WM deficits do not cause SLI, but often can contribute to it.

Deficits have also been found in the more complex WM abilities of children with SLI, which includes CE functioning. For example, Briscoe and Rankin (2009) found deficits in both simple and complex WM tasks. Seven children with SLI (ages 7;2 to 9;8) were compared to 28 children with TD, matched on nonverbal cognitive ability as well as either age or language ability. They were given a battery of WM tasks based on the tripartite model that measured nonverbal abilities, the CE, and verbal WM, as well as lexical-semantic knowledge. The results showed that language and complex WM abilities were delayed relative to same-age peers. In fact, verbal memory was more powerful than language scores at differentiating groups. Visuospatial WM was intact, however, which suggests that the WM deficit is not domain general. The authors concluded that difficulties with the phonological loop affect all levels of the tripartite model, a hypothesis of a hierarchical influence between components that contradicts the double memory jeopardy hypothesis proposed by Archibald and Gathercole (2006).

A study by Hoffman and Gillam (2004) suggested that complex CE functioning involves both the verbal and visual domains. They designed a study that intended to isolate the influence of storage limitations as well as CE limitations in both the visual and verbal modalities in school-aged children (mean age 9;5 years), 24 children with SLI and 24 children with TD. The SLI group had poorer storage and recall in both verbal and visuospatial domains in either STM tasks or complex tasks meant to tap the CE in both modalities. These results suggested more general limitations in information processing capacity in children with SLI.

In summary, evidence has accumulated to suggest that children with SLI have poorer verbal WM capacity, processing efficiency, and CE functioning (including updating), yet the mechanisms remain unclear (Gillam & Hoffman, 2004). Not all children exhibit difficulties in all of these areas, and the involvement of visual deficits is undecided. More research is thus needed to reach a better understanding of how specific mechanisms are involved, especially regarding the recently emerging topic of updating.

There is currently little data regarding updating abilities in children with SLI. In the only existing study to date, Im-Bolter and colleagues (2006) found that scores on a nonverbal updating task (n-back, see Method section below for further details) were lower for the children with SLI than for age-matched 7- to 12-year-old peers with TD. For the results of a path analysis, the authors concluded that poorer updating mediated the relationship between processing capacity and language competence in the SLI group. Participants with SLI had highly variable performance between individuals and tasks, however. Results and interpretations were also hampered by both ceiling effects (on the 0-back) and floor effects (on the 2-back) for the various conditions of the updating task. The purpose of our study is to contribute to the limited existing research on updating abilities in children with SLI, and how these abilities may relate to a particular language production ability – character referencing in narratives.

## 1.6 The present study: Research questions

The current study investigated updating and referencing abilities in children with SLI by building on Whitely and Colozzo's (under review) findings regarding these abilities in TD children, as well as on existing research pertaining to children with SLI. Both linguistic and cognitive processing factors contributing to referencing in narratives were considered. The referencing and updating abilities of the children with SLI were compared to the existing data of same-age typical peers. This study included both between-group (SLI versus TD) analyses and within-group analyses that focused on individual differences. The following research questions were considered:

1. How do children with SLI perform on updating tasks in comparison to TD peers?

   Based on the findings of Im-Bolter and colleagues (2006) and on the accumulated evidence regarding performance on WM tasks deemed to be taxing the CE (Montgomery et al., 2010), we expected the children with SLI to perform significantly worse than same-age TD peers on the various updating tasks, and that these differences would be greater for the verbal updating tasks.

2. Are referential adequacy levels similar between the SLI and TD groups?

   Based on prior research, we predicted that children with SLI would present with lower levels of referential adequacy rates compared to TD peers (Finestack et al., 2006; Liles, 1985; Norbury & Bishop, 2003; Schneider & Hayward, 2010).

3. Do the difficulties that children with SLI and TD have in referential adequacy vary systematically depending on referential function? Do the same patterns of difficulty emerge for both groups? Whitely and Colozzo (under review) found that maintenance was the easiest referential function and that reintroduction was the most challenging for TD 5- to 8-year olds. Although very little research involving children with SLI has compared the referential functions, we expected a similar pattern to emerge for these children. One might expect children with SLI to struggle even more with reintroduction given a more limited updating capacity. Alternatively, linguistic difficulties could make maintenance challenging if children are having difficulties with pronouns.

4. Are distributions of referential types similar between the SLI and TD groups? How does this interact with adequacy? The limited existing research suggests that both groups of children should present with similar patterns of use for the various referential types (i.e., nominals vs. pronominals/ellipsis) – although there could be differences depending on the referential function (e.g., fewer pronouns for maintenance). Regardless of whether differences are revealed in terms of distributions, the group with SLI might show more difficulty with both types of forms, or with only one: pronominal forms due to errors regarding gender or number; nominals as a result of difficulties with the indefinite/definite distinction that may persist longer compared to TD children. It was, however, unclear what to expect given the relatively sparse research on referencing in SLI.
5. Are individual differences in WM updating abilities related to referential adequacy in each group? Does this depend on the referential function? Based on the findings of Whitely and Colozzo (under review) for TD children, we hypothesized that a relationship would be found between updating and referencing in TD children as well as those with SLI, although the strength of these relationships and the specific factors involved could vary by group. Also, we expected the maintenance function to show the strongest and most reliable relationship with updating for children in this age range.

# 2 Method

To investigate WM updating and referencing ability in children, participants completed four updating tasks, two story-telling tasks, and a pretest to explore their production abilities regarding referential forms. The measures of updating tapped both verbal and visual WM. STM tasks were also included to ascertain to what extent differences in performance on the updating tasks may be influenced by differences in non-executive factors. The Behavioural Research Ethics Board of the University of British Columbia reviewed and approved all aspects of this study. This chapter describes participant characteristics, task and scoring procedures, and reliability.

## 2.1 Participants

Following some piloting to determine feasibility, 12 children with significant language deficits aged 5 to 8 years participated in the study. This age range was chosen to match the chronological age of the TD group from Whitely and Colozzo (under review), and was based on developmental data pertaining to both referential abilities (e.g., Beliavsky, 2003; Liles, 1993; Norbury & Bishop, 2003) and updating (e.g., Gathercole, 1998; Im-Bolter et al., 2006) in both TD children and those with SLI. Participants were recruited through speech-language pathologists (SLPs) working either in private practice or at the Vancouver School Board. The private practice and school SLPs identified children on their caseloads who met the inclusion criteria, and obtained initial consent from the parents. The parents then gave informed consent for their children to participate after having been provided detailed information about the study by the researchers.

In order to meet inclusion criteria, the participants were required to present with significant language difficulties in the context of normal nonverbal cognition, and to be monolingual English speakers. (See section 2.2.1 below for additional details regarding procedures and criteria for language and nonverbal cognition.) Monolingual English speaker was operationalized as speaking English as the primary language at home and at school, and having English as a native language. Consistent with criteria from the EpiSLI model of Tomblin and colleagues (1996), a child was not included if he/she had motor, visual, hearing, oral-structural, cognitive, or gross neurological impairments. A parent questionnaire was included with the consent forms in order to confirm that these inclusion/exclusion criteria were met. This information was also augmented by verbal reports from teachers/SLPs. Parents were also asked whether they had any concerns regarding their child's hearing and whether this had previously been tested, and if a history of ear infections existed. In addition, participants were monitored for signs of congestion at each testing visit. A hearing screening was performed at the beginning of the second session – the testing session that included tasks that particularly involved listening (see section 2.2.1 below for additional details regarding hearing screening procedures). Every participant passed the screening. One child's hearing was not screened, but there were no concerns based on reports and observations.

A norm-referenced language assessment and a nonverbal cognitive screen were used to confirm that the children with SLI had a primary language difficulty, as well as to document their levels of verbal and nonverbal abilities (see the following section for details). This study includes data for the 12 children who met all inclusion criteria for the SLI group: 4 children in kindergarten (3 boys, 1 girl; mean age 5.77 years), 4 children in grade 1 (3 boys, 1 girl; mean age 7.02 years), and 4 children in grade 2 (3 boys, 1 girl; mean age 7.71 years). These children came from various socioeconomic backgrounds, with three mothers (25%) reporting grade 11 education, one (8%) reporting high school education, four (33%) having received some additional schooling beyond high school, and four (33%) being university-educated.

The 12 participants of this study were matched with 12 participants from the sample of TD children from Whitely and Colozzo (under review) for gender and grade. When more than one child met these criteria, we selected the one who matched most closely on age and maternal education as well (see Table 1 for the means and ranges; Table 13 in Appendix C presents the age of each participant). Age was matched within 6 months, with an average difference of 1.44 months. Maternal education was matched within 2 years, with an average difference of 0.58 years. Paired sample t-tests were performed to confirm that the groups were well matched along these dimensions. As expected, there were no significant differences between groups either for age, $t(11) = 1.88$, $p = .087$; $d = .129$, or for maternal education, $t(11) = 1.47$, $p = .171$; $d = .259$.

**Table 1 Demographic data means (and ranges), by group**

| | Group | |
|---|---|---|
| | SLI ($n$ = 12) | TD ($n$ = 12) |
| Age (years) | 6.83 (5.50-8.01) | 6.95 (5.58-8.17) |
| Maternal education (years of schooling) | 13.92 (11-18) | 14.50 (12-18) |

## 2.2 Experimental tasks, procedures, and scoring

Participants completed twelve tasks in total. These tasks included: (a) two standardized tests, (b) a referential devices pretest, (c) a hearing screening, (d) four tasks to evaluate updating, (e) two STM tasks, and (f) two narrative generation tasks. All of the experimental tasks were administered over three testing sessions of about 45 minutes each. The order of the tasks was the same for each participant so that all children were equally affected by any possible order effects. Visual memory tasks occurred before verbal memory tasks to avoid leading the children to use verbal strategies on tasks intended to be visual. The narratives were collected during the final experimental session so that the children would be most comfortable with the experimenter. The first session consisted of the standardized tests (TOLD-P:4 and KBIT-2) and visual span. The second session began with the hearing screening, followed by digit span, 1-back, the referential devices pretest, and sound monitoring. The third session opened with keep track, followed by *April Fools* and *Frog on His Own*, and concluding with 2-back. The sessions were typically one week apart, although this varied due to scheduling constraints (ranging from one day to three weeks). Participants were tested individually during the last four months of the school year. For each task, familiarization and practice phases preceded the test phase. The tasks followed essentially identical procedures as those implemented successfully with children of the same age by Whitely and Colozzo (under review), who had adapted the tasks from existing research. Refer to Appendix G for detailed instructions for each task.

### 2.2.1 Language and cognitive testing

Two standardized tests were first administered to confirm the presence of SLI (see Table 2). Table 11 in Appendix A presents the standard test scores obtained by each child in the SLI group. The *Test of Language Development Primary:4* (TOLD-P:4; Newcomer & Hammill, 2008) was administered following standard scoring guidelines to assess the participants' language abilities. This test has been frequently used in research (e.g., Gillam et al., 2008; Tomblin et al., 1996) and clinical settings. It includes six core subtests that assess comprehension and production in vocabulary, grammar, and syntax. Children were considered to have a language impairment if they obtained a score of at least 1 standard deviation (SD) below the mean for at least two of the six subtests. Two children were excluded from the current study because they did not meet this criterion (i.e., their scores on the TOLD-P:4 were too high).

The *Matrices subtest* from the *Kaufman Brief Intelligence Test – 2nd Edition* (KBIT-2; Kaufman & Kaufman, 2004) assesses nonverbal cognitive abilities. It was used to confirm parent and professional reports and to describe our sample. This subtest, which assesses a child's ability to understand relationships and to complete visual analogies, has been used as a screening tool for children with SLI in prior research (e.g., Gillam et al., 2008). The participants' standard scores were obtained as indicated in the test's manual. Many small- and large-scale studies have found that school-aged children with SLI tend to score in the normal range but perform slightly lower than age-matched typical peers on tests of nonverbal cognition (e.g., Colozzo et al., 2011; Fey et al., 2004). These findings are being progressively acknowledged as likely reflecting a true difference between the groups. In accord with recent research, we used an inclusion criterion of no lower than 1.67 standard deviations below the mean, or a standard score ≥ 75 (Gillam et al., 2008). No participants obtained scores that fell in the range of intellectual disability (i.e., a standard score below or equal to 70; American Psychiatric Association, 2000), and all children met this inclusion criterion.

**Table 2 Standard scores (and standard deviations and ranges) for the SLI group (*n* = 12)**

| Variable | Mean | SD | Range |
|---|---|---|---|
| TOLD-P:4, Picture Vocabulary | 9.67 | 3.17 | 4-14 |
| TOLD-P:4, Relational Vocabulary | 6.17 | 2.98 | 2-11 |
| TOLD-P:4, Oral Vocabulary | 4.50 | 2.39 | 1-7 |
| TOLD-P:4, Syntactic Understanding | 8.83 | 2.48 | 4-13 |
| TOLD-P:4, Sentence Imitation | 5.67 | 3.08 | 1-13 |
| TOLD-P:4, Morphological Completion | 6.17 | 2.41 | 3-9 |
| KBIT-2 | 99.75 | 16.18 | 75-126 |

A *referential devices pretest* was administered to obtain data regarding the production abilities of the children with SLI at the sentence level regarding linguistic forms typically used for referencing. This pretest was meant to provide complementary information regarding which referential devices the children with SLI were able to produce, either on their own or following a model (i.e., elicited production). This data could help to ascertain whether a participant's errors occurred because he/she was not properly keeping track of the discourse model and of characters, or simply because he/she had not acquired certain linguistic forms. The probes for this pretest were selected from the linguistic devices (i.e., pronouns, determiners, relative clauses) frequently used by TD children to refer to characters in the narratives gathered by Whitely and Colozzo (under review). The selected devices were also congruent with known milestones in referential form acquisition. Piloting was conducted to verify and adjust the pretest for successful implementation.

A cloze task using picture stimuli was designed, with up to two sentences used to elicit each form (refer to Appendix H for the probe sentences for each referential device in the pretest). The examiner read the first sentence to the participant while pointing to the corresponding picture, and asked the child to finish the sentence. If the child responded incorrectly or with an unexpected response, the examiner provided the child with the correct answer and then administered the second probe for that particular form. The experimenter audio-recorded the child's responses for later verification of the online scoring. The referential devices pretest was given on a different day than both the standardized language assessment (TOLD-P:4) and the narrative tasks in order to spread out the linguistically demanding tasks and thus reduce fatigue. Table 3 presents the percentage of the children with SLI who correctly produced each referential device that was probed for. Also, Table 12 in Appendix B presents the performance by each child in the SLI group on the target items.

**Table 3 Percentage of children with SLI who correctly produced each target in the referential devices pretest**

| My | 's | Their | His | Hers | It | Them | Him | Her | He | She | I | Me | They | You | A | The | Who /that |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 73 | 82 | 64 | 82 | 73 | 91 | 54 | 54 | 91 | 82 | 64 | 91 | 91 | 36 | 91 | 100 | 100 | 0 |

*Note*. For all targets, $n = 11$.

Some of the probes might not have been designed in a way to elicit the specific linguistic form, so this pretest cannot definitively reveal which referential devices the children had not yet acquired. Nevertheless, the pretest did confirm the forms that they could produce. If there was evidence that a specific form was in a child's expressive linguistic repertoire, then errors or omissions using that form as a referential device in the narratives were more likely to have occurred because of the processing demands of character reference in narratives. Devices with lower acquisition rates included plural pronouns (*them, their, they*), feminine pronouns (*she, hers*), and one masculine pronoun (*him*). This suggests that problems referring to groups of people or individuals might be partially traced to difficulties with these pronominal forms. Also, the pronoun *who* (or *that*) was not elicited from any participant. This could either be explained by the fact that no children had acquired this form, or that the prompt was simply inappropriate to elicit it. The reason is relatively inconsequential because there are other ways for a child to get around expressing such meanings. Finally, it is important to note that both the indefinite (*a*) and definite (*the*) determiners were correctly elicited from all children. This suggests that errors in making the indefinite/definite distinction in the narratives were not attributable to an absence of either of these forms in the productive language of the children in the sample.

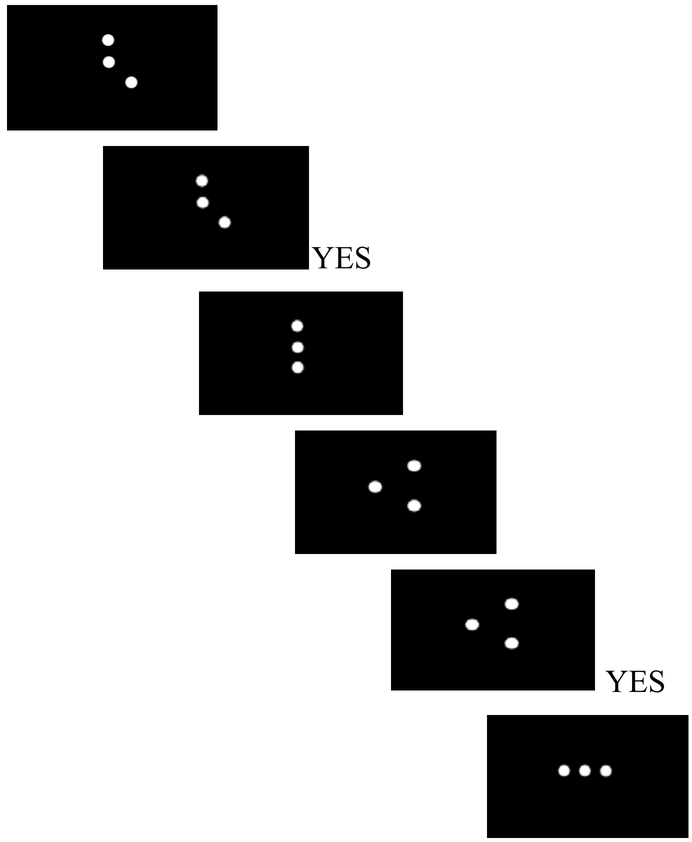
A *hearing screening* was conducted at the start of the second day of testing to ensure that hearing problems would not interfere with the tasks that involved listening. Participants indicated that they detected a series of pure tones presented through headphones. Current American Speech-Language-Hearing Association (ASHA, 1997) screening guidelines for children ages 5 and up were followed. To pass the screening, the participant had to detect tones of 1000, 2000, and 4000 Hz at an intensity level of 20 dB HL in both ears.

### 2.2.2 Memory tasks (updating and STM)

The tasks for assessing updating abilities in WM were identical to those used by Whitely and Colozzo (under review): *n-back, sound monitoring,* and *keep track*. These tasks had been adapted from previous research to be suitable for 5- to 8-year-old children. They varied in terms of presentation modality (verbal, auditory, or visual), response type (verbal or nonverbal), and the amount of verbal mediation they invited.

The *n-back task* (Im-Bolter et al., 2006; Mantyla et al., 2007) is an updating task that is commonly used in functional brain imaging studies (e.g., Owen, McMillan, Laird, & Bullmore, 2005). In the current study, this task required identity-monitoring of visual stimuli and it was presented in two conditions: 1-back and 2-back. Both conditions involved visual presentation and nonverbal responses, and did not invite verbal mediation. E-Prime software was used to program and administer the task (Schneider, Eschman, & Zuccolotto, 2002). A changing configuration of three white dots on a black background was presented in the middle of a computer screen. The image was displayed for 2000 ms, with a blank screen appearing for 500 ms between presentations. The child pressed the spacebar when he/she judged that the array was the same as a previously presented configuration. Using this simple response, Whitely and Colozzo (under review) were able to successfully avoid floor effects (which had been an issue in Im-Bolter et al., 2006) with children aged 5 to 8 years. The task required the participant to hold one or two visual representations at a time in WM for comparison to one that would appear either one (1-back) or two (2-back) images later. Each target image had to continually be replaced with a newer, more relevant image in the participant's WM. See Figure 1 for an example of a 1-back trial.

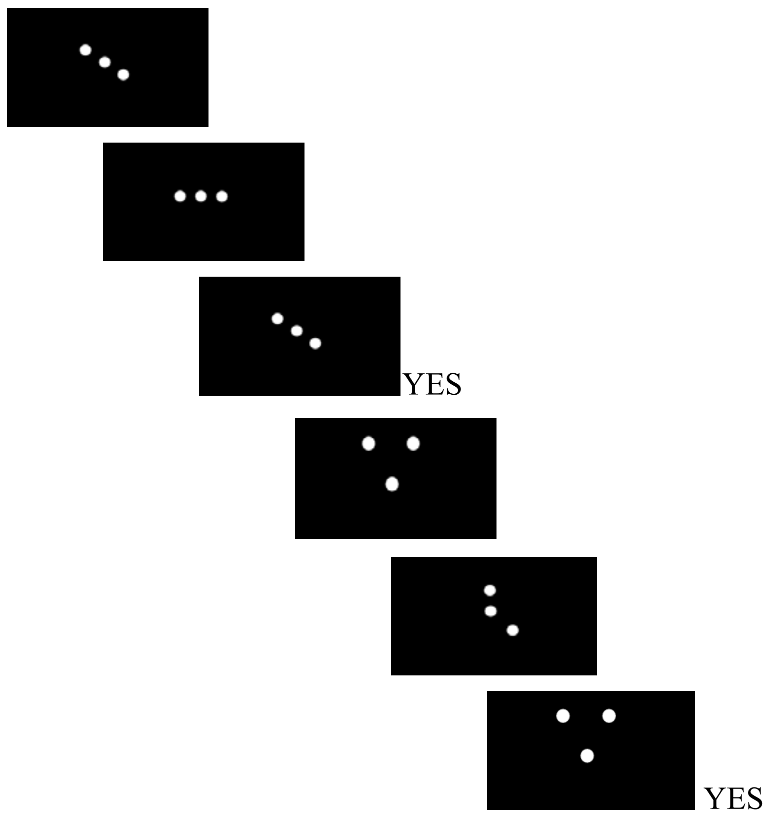
**Figure 1  Partial example of a 1-back trial**




The 1-back task and 2-back tasks were administered on the second and third days of testing, respectively. In both cases, a 6-item practice trial (using manipulable cards) was first presented to teach the task rules. Two of these images were hits, meaning that they were identical to the target. Second, the children watched a demonstration trial on the computer (nine images with three hits). The participant repeated this set as a practice trial, and if he/she did not complete it successfully, a final practice trial was administered in which the examiner did the practice with the participant one more time to provide help if errors persisted. Three test trials were presented, with 18 images shown in each, for a total of 54 items. Six or seven of the images in each trial were possible hits, with 20 possible hits and 34 correct rejections in total. E-Prime automatically recorded the participants' responses.

In the 2-back task, the child responded when the array was the same as the configuration presented two images prior (see Figure 2 for an example). This condition involved a greater degree of difficulty because two different visual representations had to be held in WM at a time. Responses were converted to hit and false alarm rates for the 1-back and 2-back conditions separately. (See section 3.1 below for additional details regarding transformations prior to statistical analyses.)

**Figure 2  Partial example of a 2-back trial**



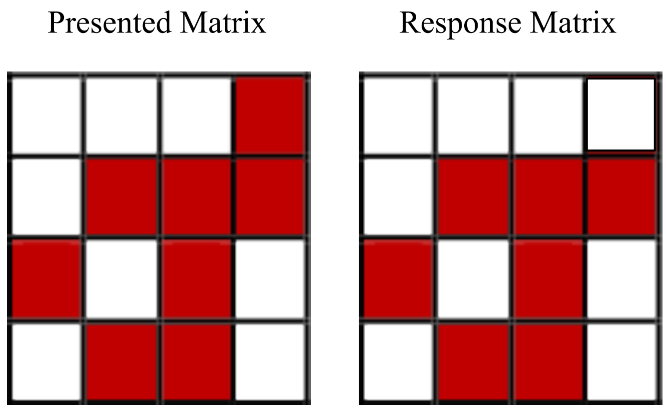

The *sound monitoring task* was adapted by Whitely & Colozzo (under review) to be appropriate for young children from a tone monitoring task (Miyake et al., 2000). This recognition task involved auditory presentation and a nonverbal response. The nature of the task made it very difficult to verbally rehearse, at least for young children. E-Prime software was used to present the task and to automatically record the participants' responses. Three different familiar environmental sounds (bird, bell, car horn) that were never labeled with words by the experimenter were presented in a random sequence three times each, for a total of nine sounds per trial. Each participant completed three trials (for a total of 9 possible hits and 18 possible correct rejections) in the same order. The sounds were presented over headphones for 500 ms each, with 1500 ms of silence between sounds. The child pressed the spacebar every time he/she detected each sound for the third time. A biological check was performed before placing the headphones on participants to ensure that they were functioning and were set to a comfortable intensity level. The experimenter first completed a demonstration trial and then the same trial was repeated as a practice trial, which was administered up to two times as needed (i.e., if the child continued to make errors). Responses were recorded online by the experimenter for comparison to the E-Prime electronic data. Hit and false alarm rates were computed for each participant. (See section 3.1 below for additional details regarding transformations prior to statistical analyses.)

The *keep track task* used in this study was developed by Yntema (1963), adapted by Miyake and colleagues (2000), and then modified further by Whitely and Colozzo (under review) to make it appropriate for use with young children. Adaptations included: changing the categories and objects of the original task to stimuli with labels acquired early in development; decreasing the number of target categories; and using only 1-2 syllable words. This recall task required categorization and labelling. It involved visual presentation and a verbal response, and had strong potential for verbal mediation because the pictures were familiar and thus easily labelled. A series of colour pictures of familiar objects from six categories (food, animals, body parts, toys, clothing, and transportation) were shown on the computer screen (see Appendix I for the list of all objects used in this task). Eight pictures were included for each of the six categories, for a total of 48 pictures. Throughout each trial, icons that represented each of the (two or three) target categories were shown on the bottom of the screen as a reminder. The pictures were presented one at a time in random order for 2000 ms, with 1000 ms of blank screen between them. When a trial finished, a question mark was displayed on the screen and the child was asked to verbally recall the last item presented for each target category. The task involved three levels of difficulty. Each level included three trials, creating a total of nine trials completed in the same order by every participant. Trials in Condition 1 consisted of two target categories in each of three 10-picture trials. Each trial included three pictures from each target category and four foils from non-target categories. Inclusion of foils was meant to discourage participants from only attending to pictures near the end of the presentation. Trials in Condition 2 consisted of two target categories and 12 pictures, including four pictures from each target category and four foils. Finally, trials in Condition 3 consisted of three target categories in each 12-picture trial, including three pictures from each target category and three foils. The trials were balanced in terms of how frequently each category and picture came up, as well as syllable length, so that the order of the pictures was quasi-random. Prior to the experimental trials, each participant was shown icons for each of the six categories. To ensure familiarity with each picture and its category, each participant labeled all 48 pictures and matched them to their corresponding category icon by pointing to it. The experimenter provided the correct answer if a participant made an error in labelling or categorizing a picture. A demonstration trial (two target categories and eight pictures) was then performed by the experimenter. The participant then completed a practice trial with the same stimuli, and the experimenter helped the child to complete the practice once as needed. The child's responses were recorded online by the experimenter who allotted one point for each item recalled correctly (maximum 21 points).

Two STM tasks were also included: *forward digit span* and *visual span*. Both tasks are commonly used for STM assessment in children (Gathercole, 1999). The *forward digit span task* was the Number Repetition Forward subtest task from the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003), which was administered according to the standardized procedure. The stimuli were presented verbally and a verbal response was required. It was therefore the STM counterpart of the keep track task. A sequence of digits (from 1 to 9) was presented verbally in random sequence, and the child tried to repeat each series in the same order. The first level included two trials of two-digit strings. If the participant produced every digit in the correct order on at least one trial at that level, the next level (with sequences increased by one digit per level) was administered. The highest level involved nine-digit sequences. When the child was unsuccessful for both trials within a level, the task was discontinued. Each participant obtained a score corresponding to the total number of trials he/she correctly recalled (maximum 16 trials). This task was audio-recorded to later verify online scoring.

The *visual span task* corresponded to the visual immediate memory task created by Logie and Pearson (1997). This task was the STM counterpart of the n-back task, since they are both recognition tasks involving visual presentation and nonverbal responses. A series of matrices of squares were presented for 2 seconds on the computer screen, with half of the cells coloured red (see Figure 3 for an example). The pattern disappeared to a blank screen for 2 seconds, and then the matrix reappeared with one of the previously red cells now white. The pattern remained on the screen until the child pointed to the cell he/she thought had changed. The distribution of red squares and the square that changed colour was randomly determined. Every participant completed the trials in the same order. A 2 x 2 matrix was presented as a practice trial. If the child's response was incorrect, the examiner modeled the correct response and the child repeated the practice trial. The number of cells gradually increased until the child failed to answer at least 2 of 3 trials correctly at one level, at which point the task was discontinued. The size of the matrix increased by two squares (one red, one white) at each successive level up to a maximum of 34 squares. The additional squares were added to form a new row, then a new column, and so on. The examiner recorded responses online. Each participant's score corresponded to the total number of arrays where the child responded correctly (i.e., trials correct; maximum 48 trials).

**Figure 3 Example of a trial from the visual span task**




### 2.2.3 Scoring of the memory tasks

For the keep track task and both STM tasks, each child received a score corresponding to the total number of correct items or trials, and these scores were used for statistical analyses. For the n-back and sound monitoring tasks, each participant's responses were converted to a proportion correct score that considered *hits* (responding 'yes' when the image was identical to the target) as well as *correct rejections* (no response when not identical), *p(c)** = (hits + correct rejections)/total. For the statistical analyses, d-prime (*d'*) was used to represent each child's accuracy based on signal detection theory. This value provides a measure that is unaffected by response bias (i.e., the participant's tendency toward 'yes' or 'no' responses) and that allows for comparison of sensitivity across tasks (MacMillan & Creelman, 2005). In order to obtain *d'* values, each child's *hit rate* (*h*) and *false alarm rate* (*f*; 1- correct rejection rate) were calculated separately for the 1-back and 2-back tasks. The hit rate is the proportion of accurate detections: the ratio of 'yes' responses when the item and the target are identical. The false alarm rate is the proportion of false positives: the ratio of 'yes' responses when the item and the target do not match. These proportions were then converted to *z*-scores using the inverse of the standard normal cumulative distribution in order to calculate *d'* = z(*h*) –z(*f*) for each task.[2]

### 2.2.4 Narrative production

Two wordless picture books were used to elicit the narrative productions: *April Fools* by Fernando Krahn (1974), and *Frog on His Own* by Mercer Mayer (1973). There is a long history of effectively using wordless pictures books to elicit narratives with children within this age range. The books are comparable in terms of visual appearance, length, and complexity. Different challenges in referencing, however, are evident from each book. In *April Fools*, two boys are the main protagonists who build a monster to scare many individuals and groups of people around town. In *Frog on His Own*, a frog is the main protagonist, with most of the other characters being humans of both genders. A pet frog goes on an adventure in the park, causing trouble along the way until his owner finds him (refer to Appendix F for plot details of each story). There are more characters in *Frog On His Own* (13 characters) compared to *April Fools* (6 characters). Gathering two separate narratives created more tokens for data analysis. This provided the children with a greater opportunity to use referencing, perhaps creating a more representative sample of their referencing abilities.

Starting with *April Fools*, participants looked through the pages to familiarize themselves with the story. The participant was asked to tell a story from the pictures when he/she was ready. In order to decrease the amount of shared knowledge between the narrator and the listener, the child was told that the story would be recorded for another person who would listen to it later without having the book. The experimenter also sat far enough away so that she was unable to see the pictures, thus decreasing the amount of shared contextual information. The same

[2] Adjustments were required in order to calculate *d'* in cases of either a perfect hit rate ($h = 1$) or a null false alarm rate ($f = 0$). In line with usual practice (Macmillan & Creelman, 2005) and in order to insure symmetry, these extreme hit and false alarm rates were adjusted based on the total number of responses: for $h = 1$: $1 – (1/N)$; for $f = 0$: $1/N$.

procedure was repeated for *Frog on His Own*, with a short break provided between both narrative production tasks, if necessary. Both narratives were audio- and video-recorded for later transcription and coding, as well as to note any extralinguistic cues the children used to refer to characters.

### 2.2.5 Transcription and coding of the narratives

The two narrative productions were transcribed from the recordings using the Systematic Analysis of Language Transcripts software (SALT; Miller & Inglesias, 2006) and according to SALT transcription conventions. Stories were segmented into communication units (C-units) according to Loban's (1976) criteria of a C-unit containing a main clause plus any dependent phrases or clauses. Coordinated clauses (e.g., using *and, or, but*) were segmented into separate C-units unless the same subject in both clauses was omitted in the second clause. The same rules were followed to segment multiple clauses in direct quotations. Unintelligible, abandoned, or interrupted utterances were excluded from analyses. Mazes (e.g., false starts, repetitions, filler words, and retraces), tangential comments, simple story closings (e.g., "the end"), questions to the examiner, and responses to requests for clarification were also excluded from analyses. A second coder checked all transcriptions and any discrepancies were discussed and amended.

Each first and subsequent mention of a story character was then coded and analyzed on three dimensions: referential adequacy, referential function, and referential type (see Appendix E for detailed narrative coding decisions). First, *Referential adequacy* corresponds to whether the child used each linguistic device correctly. A reference is considered adequate if the listener is able to clearly ascertain which story character the speaker referred to. This study followed a classification system adapted by Whitely and Colozzo (under review) from Halliday and Hasan (1976), Liles (1985), and Beliavsky (2003):

Adequate Forms:

A. Complete reference: This indicates that the listener is able to understand the intended identity of the referent without difficulty or ambiguity based on the linguistic form. It can refer to a character mentioned either before (anaphoric) or after (cataphoric) the current referent. Complete reference serves as a cohesive device by clearly connecting various parts of a narrative.

Inadequate Forms:

A. Incomplete reference: This denotes that the linguistic device refers to something the speaker failed to specify or to provide sufficient information about in the narrative. For example, introducing a character with a pronoun or a definite article was considered an incomplete reference.

B. Ambiguous reference: This involves using an unclear linguistic device that could possibly refer to more than one character in the text, thereby confusing the listener about the character's identity. Examples of ambiguous referencing include a pronoun that can refer to more than one character, an indefinite article used for reintroduction, errors in lexical choice (e.g., switching from *he* to *it*; changing the name of a character), and inappropriate ellipsis.

C. Exophoric reference: This occurs when the referent's identity is not provided within the text of a narrative. Instead, the speaker refers to a character exclusively through the context and information outside of the text, such as extralinguistic cues (e.g., pointing to the picture).

Second, *Referential function* refers to the child's purpose for using each linguistic device. The following classification system was adapted by Wong and Johnston (2004) from the work of Bamberg (1987):

A. Introduction: The device is used to mention a character for the first time.
B. Maintenance: The device is used to repeatedly refer to the same character within one utterance or in successive utterances. Note, however, that an utterance is not intervening if it does not advance the story (such as a descriptive statement with no reference to any characters) or if it consists of dialogue, so that characters can be maintained across such utterances.
C. Reintroduction: The device is used to link back to a previously introduced character, with intervening utterances between mentions of the same character; during this interval, attention had temporarily shifted to different characters or to different story elements.

Third, *Referential type* defines the kind of linguistic device used to refer to a character in the stories. This study used four categories based on a classification system created by Halliday and Hasan (1976), and modified by Whitely and Colozzo (under review):

A. Personals: Personal pronouns (e.g., *he, she, they, it*).
B. Possessives: Possessive pronouns and determiners (e.g., *his, hers, their*).
C. Nominals: Proper names, as well as nouns combined with definite articles, indefinite articles, and demonstratives (e.g., *Joe, a dog, the frog, this boy*).
D. Ellipsis: Omission of a noun phrase.

This coding system resulted in a total number of references to story characters, which were then considered along the dimensions mentioned above: Adequacy (Adequate or Inadequate), Function (Introduction, Maintenance, Reintroduction), and Type (Personals, Possessives, Nominals, Ellipsis). To control for differences in numbers of tokens, these values were converted to proportions. Before statistical analyses, all proportions were converted using the arcsine transformation. It is generally recommended to use the arcsine transformation for inferential statistical analyses because the means and variances of proportional data are correlated, which makes them less suitable for analysis (Winer, 1991). This transformation also serves to create greater dispersion in the tails of the distributions of scores (Judd, McClelland, & Ryan, 2009).

## 2.3 Reliability

### 2.3.1 Sample of children with SLI

For the n-back and sound monitoring tasks, online scoring was confirmed using the E-Prime electronic scoring record, with the latter given precedence in cases of disagreement.[3] For the referential devices pretest, the children's responses were audio-recorded in order to later verify online scoring.

A second coder checked all transcripts for transcription accuracy, utterance segmentation, and coding decisions regarding referencing. Any discrepancies were discussed and resolved. The stories of two children (17% of the sample) were independently transcribed and segmented into C-units by a native English speaker. Agreement was 92% for word-for-word transcription and 95% for utterance segmentation into C-units. Reliability of referential coding was estimated at 98% for character, 98% for function, and 95% for adequacy based on the number of coding changes in the samples of two participants. Finally, the transcripts of two children were independently coded for referential type: reliability was 98%.

[3] For the remaining memory tasks (keep track, visual span, and digit span), the online scoring was assumed to be exact. This decision was based on the fact that Whitely and Colozzo (under review) used video or audio recordings to assess the reliability of the online scoring for identical versions of these same tasks based on 10% of the sample and found reliability to be 100%.

### 2.3.2 Sample of TD control children

A graduate research assistant independently transcribed and segmented the stories of 7 (11%) of the 64 TD children in the sample from which our 12 matched TD participants were drawn (Whitely and Colozzo, under review). Agreement was 95% for transcription (at the word level) and 94% for utterance segmentation. An estimate of inter-rater reliability for referential coding was obtained based on the discrepancies between the two coders for the stories of 7 children (11% of the sample). Agreement was estimated at 98% for the presence of a character reference (i.e., accounting for omissions by one of the coders). For references identified by both coders, agreement was 99% for character, 99% for referential function, and 96% for referential adequacy. The range of reliability across the children and all the variables coded was 94-100%. Finally, the character references in the stories of 7 children (11% of the sample) were coded independently for referential type: agreement was high at 99%.

# 3 Results

This chapter presents the results in the order that they apply to each research question. First, performance by each group is summarized for all memory tasks and for narrative production, with a particular focus on referential adequacy. Next, the relationships between referential adequacy and both referential type and referential function are described. Finally, analyses of the relationships between updating abilities and referential adequacy are presented.

## 3.1 Memory tasks performance

The first research question asked how children with SLI would perform on updating tasks in comparison to TD peers. Two STM tasks were also included to explore possible mediating effects of STM on updating performance. All children completed the five memory tasks except for one child with SLI, for whom data for visual span performance is missing because of an oversight on the experimenter's behalf. All but one child clearly indicated that they understood the task instructions. This child also responded very rarely, so all of his memory data was excluded because his responses were deemed not valid. As predicted, the group of children with SLI performed more poorly than did same-age TD peers on all memory tasks. Neither floor nor ceiling effects were generally an issue. One child with SLI and three children with TD were at ceiling for sound monitoring, and one child with TD obtained a perfect score for keep track.

In contrast to the other three updating tasks, the 2-back task was very difficult for both groups of children, with many participants obtaining scores near chance level (*d*' = 0). This parallels the findings in prior research that used this WM task with children of similar ages. Im-Bolter and colleagues (2006) found floor effects on the 2-back task in their sample of 7- to 10-year-old children. Using a procedure identical to the one in the current study, Whitely and Colozzo (under review) also found that many children with TD did poorly on the 2-back task, and this group of children with SLI obtained even lower scores. Hence, we excluded the 2-back data from our statistical analyses.

Before completing statistical analyses, all data were inspected for univariate outliers using points that were beyond 1.5 times the interquartile range from the lower or upper quartiles as criteria. Unless it is otherwise specified, no outliers were found. Table 4 shows the mean scores and standard deviations for both groups on the three updating tasks and the two STM tasks (also see Table 13 in Appendix C for the memory scores of each participant). These differences were tested using separate one-way ANOVAs between groups for each memory task and a one-tailed critical *p*-value of 0.05 using the Holm method to control for multiple comparisons (see Table 4 for the ANOVA results for each task).[4] A significant group difference was found for all memory tasks except visual STM; in all cases, the group of children with SLI performed more poorly than did same-age peers. The most verbal updating task, keep track, showed the largest difference.[5]

---

[4] The significance levels reported for these tests are one-tailed, as there was an expected direction for the relationships tested. Given the relatively small sample size, this reduces the likelihood of committing a type II error.

Throughout the study, the Holm correction was chosen to control for Type I error when appropriate. This method is more powerful yet never rejects fewer comparisons than the Bonferonni procedure (Aickin & Gensler, 1996). Consequently, it proved to be a better choice to also guard against Type II error given the relatively small sample size. First, the *p*-values for all the comparisons are placed in increasing order. Then, each *p*-value is compared with $\alpha/(Nc - i + 1)$ for rejection of the null hypothesis, where Nc corresponds to the number of comparisons, and *i* to the rank of each specific comparison. No further tests are done beyond the first non-rejection.

[5] All tests for homogeneity of variances between groups for each task were nonsignificant (i.e., $p > .05$. Two tasks nevertheless had marginal *p*-values: 1-back ($p = .097$) and sound monitoring ($p = .062$). Results were unchanged, however, when we considered instead the Brown-Forsyth ANOVAs for unequal variances. Results were also unchanged following sensitivity analysis for keep track after removing one outlier, $F(1, 20) = 7.11$, $p = .007$. Finally, the group differences in the visual span task remained nonsignificant when tested using a paired-samples samples t-test, $t(9) = .844$, $p = .421$.

**Table 4 Means (and standard deviations) for performance on the updating and STM tasks, by group, and results of the one-way ANOVAs for each task**

| Memory task | Group | | Results of ANOVAs | | |
|---|---|---|---|---|---|
| | SLI | TD | *F* | *p* | *d* |
| 1-back, | | | | | |
| *p(c)** | 0.86 (0.09) | 0.93 (0.03) | | | |
| *d'* | 2.34 (0.81) | 3.01 (0.45) | 6.31 | .010 | 1.08 |
| Keep track, | | | | | |
| *items correct* (max. 21) | 8.73 (2.57) | 13.08 (4.34) | 8.37 | .004 | 1.26 |
| Sound monitoring, | | | | | |
| *p(c)** | 0.79 (0.13) | 0.93 (0.09) | | | |
| *d'* | 1.66 (1.20) | 2.81 (0.84) | 7.18 | .007 | 1.13 |
| Visual span, | | | | | |
| *trials correct* (max. 48) | 12.80 (6.83) | 15.75 (5.67) | 1.23 | .141 | .047 |
| Digit span, | | | | | |
| *trials correct* (max. 16) | 5.27 (1.85) | 6.92 (1.38) | 5.91 | .012 | 1.02 |

*Note*. *d* = Cohen's d. For 1-back and sound monitoring, *d'* was used in all statistical analyses. For all but one task, *n* = 11 for SLI and *n* = 12 for TD. For visual span, *n* = 10 for SLI and *n* = 12 for TD.

## 3.2 Story characteristics and referential adequacy

The following subsections consider performance by the SLI and TD groups in terms of the narratives they produced. First, basic characteristics of the narratives will be described, including productivity, length, lexical diversity, and number of references. Next, referential adequacy will be investigated for group differences in overall adequacy, and for whether the referential function and the story influenced adequacy for each group. Finally, referential type will be considered in terms of how the distributions of types compare between groups, as well as whether the type of referential device used impacted the adequacy of referencing.

### 3.2.1 Productivity, length, and lexical diversity

All participants told two stories from the wordless picture books *April Fools* and *Frog on His Own*. SALT software (Miller & Inglesias, 2006) was used to calculate four descriptive indices: total number of C-units, mean length of C-units in words (MLCU-w), number of different words (NDW), and number of references. Table 5 presents the means and ranges for story length, utterance length, and lexical diversity. In general, *Frog on His Own* elicited longer stories with greater lexical diversity for both groups. Utterance length was very similar between stories for both groups. The children with SLI produced more utterances on average, but these utterances tended to have fewer words than those produced by the TD group, and this pattern held for both stories. The TD group had greater lexical diversity than the SLI group for *April Fools*, but the opposite trend appeared for *Frog on His Own*, although the two groups were more similar in the latter case. Finally, the mean total number of character references produced by the SLI group was quite similar to that of the TD group for both stories.

A 2 x 2 mixed ANOVA was performed to analyze the effects of group and story on the total number of references using a 2-tailed critical *p*-value of .05. There was no main effect of group, as both groups produced a similar number of references ($M_{SLI} = 55.3$; $M_{TD} = 53.8$), $F(1, 22) = .050$, $p = .825$; $\eta_p^2 = .002$. There was, however, a main effect of story, as *Frog on His Own* ($M = 60.0$) elicited more referential tokens compared to *April Fools* ($M = 49.0$), $F(1, 22) = 15.4$, $p = .001$; $\eta_p^2 = .412$. There was no significant interaction between the variables, $F(1, 22) = .837$, $p = .370$; $\eta_p^2 = .037$, suggesting that the stories affected the groups similarly.

**Table 5 Means (and ranges) for number of utterances (C-units), mean length of utterances in words (MLCU-w), number of different words (NDW), and number of character references, by story and by group**

| Story | | Group | |
|---|---|---|---|
| | | SLI | TD |
| April Fools | C-units | 30.2 (17-44) | 26.2 (13-50) |
| | MLCU-w | 6.08 (3.86-7.83) | 7.98 (6.85-8.69) |
| | NDW | 74.1 (29-115) | 80.0 (43-148) |
| | References | 48.4 (23-74) | 49.5 (21-85) |
| Frog on His Own | C-units | 37.7 (21-56) | 30.1 (14-39) |
| | MLCU-w | 6.27 (4.48-7.93) | 7.87 (5.72-10.26) |
| | NDW | 90.4 (51-148) | 88.1 (49-123) |
| | References | 62.1 (30-101) | 58.0 (22-78) |

### 3.2.2 The influences of group, referential function, and story on referential adequacy

The following analyses considered whether the two groups of children differed with regards to their levels of referential adequacy, and whether any such differences were influenced by referential function. Also, given the significant differences in terms of the numbers of referential tokens between the stories, we considered the possible influence of story as well.

Table 6 and Figure 4 present the mean proportions of adequate references by group and by function for each story (also refer to Table 13 in Appendix C for the referential adequacy scores of each participant, by story). The SLI group obtained lower overall referential adequacy scores than did the TD group for each of the stories. These differences were, however, not as apparent for introduction and reintroduction for *April Fools*, whereas they were consistent across functions for *Frog on His Own*. Both groups of children consistently showed the highest levels of adequacy for maintenance compared to introduction and reintroduction. They also obtained lower levels of adequacy across functions for *April Fools* compared to *Frog on His Own*.

**Table 6 Referential adequacy measured in mean proportion of complete references (with standard deviations), by function and by group, for each story**

| Story | Function | Group | | Both groups |
|---|---|---|---|---|
| | | SLI | TD | |
| April Fools | Overall | 0.57 (0.19) | 0.70 (0.11) | 0.64 (0.16) |
| | Introduction | 0.54 (0.33) | 0.55 (0.18) | 0.55 (0.26) |
| | Maintenance | 0.71 (0.16) | 0.85 (0.10) | 0.78 (0.15) |
| | Reintroduction | 0.39 (0.22) | 0.40 (0.26) | 0.39 (0.24) |
| Frog on His Own | Overall | 0.69 (0.13) | 0.82 (0.09) | 0.75 (0.13) |
| | Introduction | 0.56 (0.28) | 0.68 (0.23) | 0.62 (0.26) |
| | Maintenance | 0.79 (0.10) | 0.90 (0.10) | 0.84 (0.11) |
| | Reintroduction | 0.59 (0.17) | 0.78 (0.11) | 0.68 (0.17) |

**Figure 4a Referential adequacy, by function and by group, *April Fools***

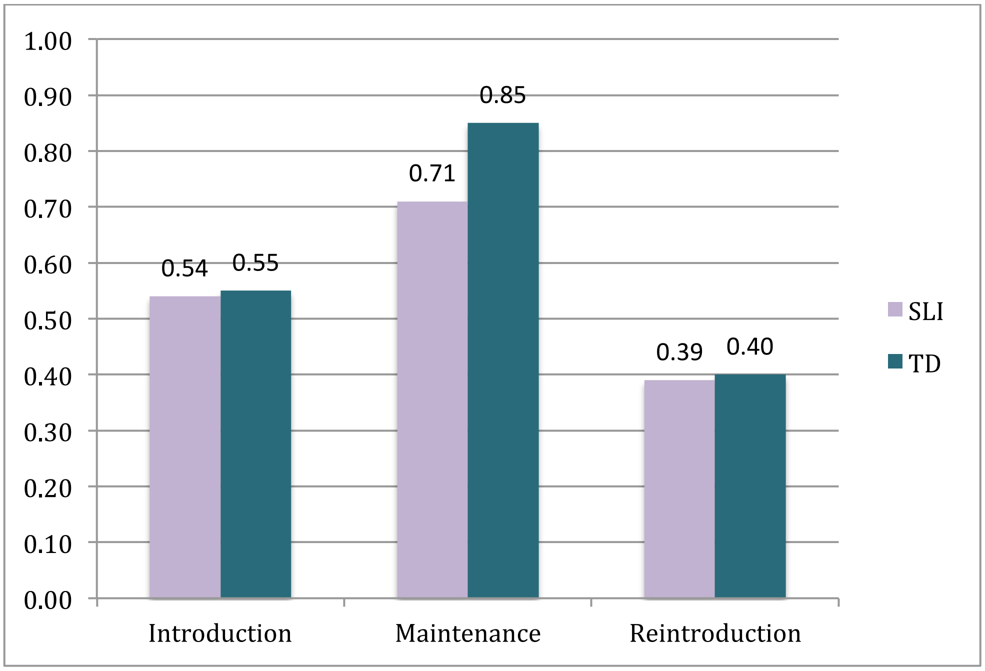
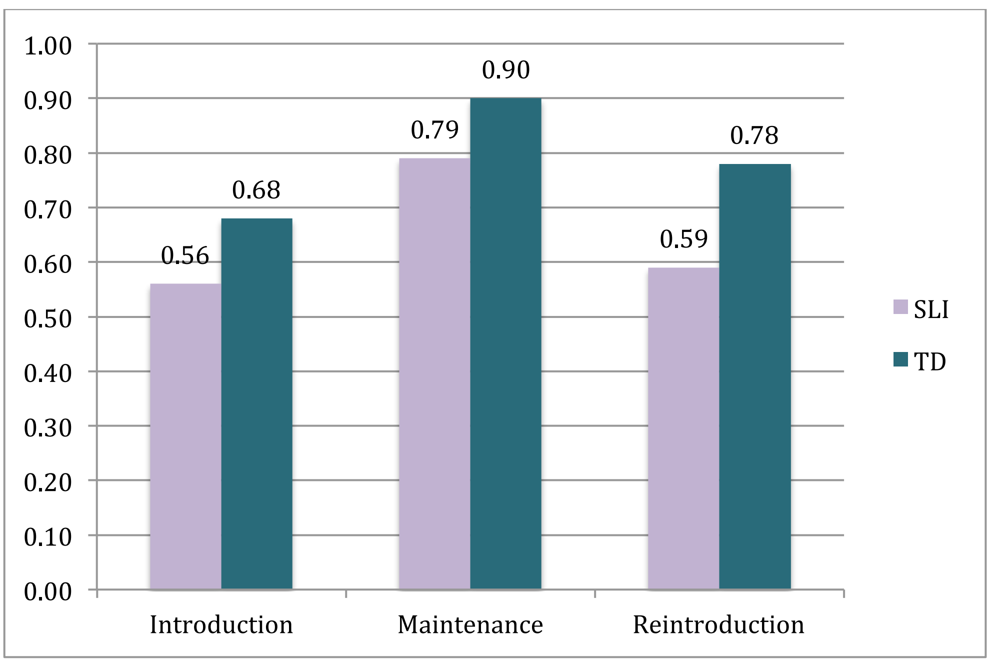



**Figure 4b Referential adequacy, by function and by group, *Frog on His Own***




To test whether these observed patterns were significant, a 2 x 3 x 2 mixed ANOVA analyzed the effects of group, function, and story on referential adequacy using the arcsine transformed values of proportion complete references. The main effect of group was not significant ($M_{SLI}$ = 1.77; $M_{TD}$ = 2.03),[6] $F(1, 22) = 3.05$, $p = .095$; $\eta_p^2 = .122$. There was, however, a significant main effect of function $F(2, 44) = 30.3$, $p < .001$; $\eta_p^2 = .579$, with maintenance having the highest adequacy level ($M = 2.30$), followed by introduction ($M = 1.74$), and by reintroduction ($M = 1.66$). Finally, there was also a main effect of story, with significantly lower overall adequacy levels for *April Fools* ($M = 1.73$) compared to *Frog on His Own* ($M = 2.07$), $F(1, 22) = 39.7$, $p < .001$; $\eta_p^2 = .643$.[7]

There were also two significant interactions: between group and story, $F(1, 22) = 4.34$, $p = .049$; $\eta_p^2 = .165$, and between function and story, $F(2, 44) = 4.95$, $p = .011$; $\eta_p^2 = .184$. Neither the function by group nor the three-way interaction were significant, with all *p*-values > .542. Hence, the two groups were similarly affected by the various referential functions.

To follow up the story by group interaction, two one-way ANOVAs were performed for overall referential adequacy by group, one for each story. The groups did not differ significantly for *April Fools*, $F(1, 22) = 3.97$, $p = .059$; $d = .813$, whereas they did differ significantly for *Frog on His Own*, $F(1, 22) = 7.73$, $p = .011$; $d = 1.13$.

To follow-up the function by story interaction, separate repeated-measures ANOVAs were performed for each story to analyze the effect of referential function on referential adequacy. For *April Fools*, there was a significant effect of function, $F(2, 46) = 22.3$, $p < .001$; $\eta_p^2 = .493$; contrasts indicated that referential adequacy was significantly higher for (i) maintenance ($M = 2.20$) compared to introduction ($M = 1.65$), $F(1, 23) = 23.4$, $p < .001$; $\eta_p^2 = .505$; (ii) for maintenance compared to reintroduction ($M = 1.34$), $F(1, 23) = 42.9$, $p < .001$; $\eta_p^2 = .651$; and, (iii) for introduction compared to reintroduction, $F(1, 23) = 4.70$, $p = .041$; $\eta_p^2 = .170$.

---

[6] The marginal means reported correspond to the arcsine transformed values for proportion complete references.

[7] One outlier was found when considering the functions combined. Results of the mixed ANOVA were unchanged when we removed the outlier.

For *Frog on His Own*, there was a significant effect of function, $F(2, 46) = 13.5$, $p < .001$; $\eta_p^2 = .370$. Contrasts once again indicated that referential adequacy was significantly higher (i) for maintenance ($M = 2.40$) compared to introduction ($M = 1.83$), $F(1, 23) = 18.3$, $p < .001$; $\eta_p^2 = .442$; and (ii) for maintenance compared to reintroduction ($M = 1.99$), $F(1, 23) = 37.5$, $p < .001$; $\eta_p^2 = .620$. For this story, however, performance was similar between introduction and reintroduction, $F(1, 23) = 1.64$, $p = .213$; $\eta_p^2 = .066$.[8]

To summarize, with regards to referential adequacy, there were both similarities and differences between groups. A significant difference in overall adequacy emerged only for *Frog on His Own*, although there was a strong trend in that direction for *April Fools* as well. The groups were comparably affected by referential function, with maintenance showing higher levels of adequacy than either of the switching functions for both stories, but the two switching functions differing from each other only for *April Fools*. The groups were also similarly affected by story, with referential adequacy proving significantly lower for *April Fools* than for *Frog on His Own*.

### 3.2.3 The influences of group, referential function, and story on referential type

The next analyses explored whether the groups of children differed with regards to the distributions of the referential types. Type was dichotomized into nominals (indefinite and definite noun phrases; proper nouns) and pronominals (all types of pronouns including possessives; ellipsis). We also considered the possible influences of referential function and story on referential type distributions due to the significant effects of these variables in previous analyses. The mean proportions of pronominal forms (including ellipsis) by group, by function,

[8] The data were checked for outliers. No outliers were found when considering referential adequacy overall for each story. When functions were considered separately for each story, however, there were four outliers in *April Fools* and two outliers in *Frog on His Own*. This is an unwanted consequence of our small sample and we chose not to remove the outliers.

and by story are presented in Table 7. There were no clear differences between the groups in proportions of pronominal forms, and the groups appeared to be influenced by function and by story in comparable ways.

A mixed 2 x 3 x 2 ANOVA was performed to analyze the effects of group, function, and story on the distributions of referential types using the arcsine transformed proportions of pronominals. The main effect of group was not significant ($M_{SLI}$ = 1.55; $M_{TD}$ = 1.57),[9] $F(1, 22)$ = .027, $p$ = .871; $\eta_p^2$ = .001, with both groups using a similar proportions of pronominal forms. As predicted, there was a significant main effect of function ($M_{intro}$ = 1.04; $M_{maint}$ = 2.18; and $M_{reintro}$ = 1.46), $F(1, 44)$ = .46.7, $p < .001$; $\eta_p^2$ = .680, with the most pronouns used for maintenance. There was also a significant effect of story, as both groups used a higher proportion of pronominal forms in *April Fools* ($M$ = 2.02) compared to *Frog on His Own* ($M$ = 1.11), $F(1, 22)$ = 86.2, $p < .001$; $\eta_p^2$ = .797. Finally, there was a significant interaction of function by story, $F(2, 44)$ = 4.15, $p$ = .022; $\eta_p^2$ = .159. All other interactions involving group were nonsignificant, with all $p$-values > .117.

[9] Here, the marginal means reported correspond to the arcsine transformed values for proportions of pronominals.

**Table 7 Mean proportions (and standard deviations) of pronominal forms, by story, by function, and by group**

| Story | Function | Group | | Both groups |
|---|---|---|---|---|
| | | SLI | TD | |
| April Fools | Overall | 0.73 (0.15) | 0.82 (0.08) | 0.77 (0.12) |
| | Introduction | 0.44 (0.40) | 0.58 (0.20) | 0.51 (0.32) |
| | Maintenance | 0.85 (0.11) | 0.91 (0.05) | 0.88 (0.09) |
| | Reintroduction | 0.63 (0.25) | 0.67 (0.20) | 0.65 (0.22) |
| Frog on His Own | Overall | 0.44 (0.15) | 0.46 (0.15) | 0.45 (0.14) |
| | Introduction | 0.15 (0.18) | 0.08 (0.15) | 0.12 (0.17) |
| | Maintenance | 0.59 (0.21) | 0.68 (0.17) | 0.63 (0.19) |
| | Reintroduction | 0.31 (0.18) | 0.21 (0.19) | 0.26 (0.19) |

*Note*. Pronominal forms include both pronouns and ellipsis.

To follow-up the interaction of function by story, two one-way ANOVAs were performed, one for each story. For *April Fools*, there was a significant effect of function, $F(2, 46) = 15.6$, $p < .001$; $\eta_p^2 = .404$. Contrasts specifically showed that maintenance ($M = 2.49$) involved a significantly higher proportion of pronominals than did introduction ($M = 1.61$), $F(1, 23) = 20.9$, $p < .001$; $\eta_p^2 = .476$, and reintroduction ($M = 1.96$), $F(1, 23) = 23.9$, $p < .001$; $\eta_p^2 = .509$. There was also a significant difference between introduction and reintroduction, $F(1, 23) = 4.68$, $p = .041$; $\eta_p^2 = .169$. For *Frog on His Own*, there was once again a significant effect of function, $F(2, 46) = 48.0$, $p < .001$; $\eta_p^2 = .676$. Contrasts specifically showed a significantly higher proportion of pronominals for maintenance ($M = 1.87$) than for introduction ($M = 0.48$), $F(1, 23) = 69.1$, $p < .001$; $\eta_p^2 = .750$, and reintroduction ($M = 0.97$), $F(1, 23) = 85.1$, $p < .001$; $\eta_p^2 = .787$. There was also a significant difference between introduction and reintroduction, $F(1, 23) = 9.58$, $p = .005$; $\eta_p^2 = .294$. Overall, the same patterns in distributions of types by function (i.e., proportion of prononimals) emerged for both stories, although the effect sizes indicate different degrees of difference between the various functions. Specifically, on the one hand, *Frog on His Own* presented with larger differences between maintenance and each of the switching functions and, on the other hand, a smaller difference between the two switching functions.

To summarize, the groups used a comparable proportion of pronominal forms overall and for the various referential functions in each of the stories. Both groups were thus likewise influenced by function—with a higher frequency of pronouns for maintenance compared to either of the switching functions. They were also similarly influenced by story, with both groups using a higher proportion of pronominals in *April Fools* compared to *Frog on His Own*.

### 3.2.4 The influence of referential type on referential adequacy

In addition to the distribution of referential types, this study considered whether the groups of children differed in the adequacy of their use of each referential type. Results so far indicated significant differences in adequacy between the groups only for *Frog on His Own* in the context of no group differences in terms of the distributions by type (nominal vs. pronominal forms) for either story. The group with SLI could nonetheless show more difficulty than TD peers with both types of forms, or with only one. The possible influences of referential function and of story on the adequacy of each referential type was once again considered. As shown in Table 8, using pronominal forms (including ellipsis) resulted in lower referential adequacy for each group and each story, although the difference was much smaller for the TD group in *Frog on His Own*. For each story, the group of children with SLI had higher percentages of both inadequate nominals and of inadequate pronominal forms compared to the group of TD peers.

To confirm these observations, a 2 x 2 x 2 mixed ANOVA was performed to analyze the effects of group, type, and story on referential adequacy. There was a significant main effect of group, with lower adequacy for the SLI group ($M$ = 1.91) compared to the TD group ($M$ = 2.24),

$F(1, 21) = 7.98$, $p = .010$; $\eta_p^2 = .275$.[10] The main effect of type was also significant, with greater referential adequacy for nominals ($M = 2.21$) than for pronominals ($M = 1.94$), $F(1, 21) = 25.6$, $p < .001$; $\eta_p^2 = .550$. There was also a significant main effect of story, with lower referential adequacy in *April Fools* ($M = 2.00$) compared to *Frog on His Own* ($M = 2.15$), $F(1, 21) = 7.45$, $p = .013$; $\eta_p^2 = .262$. None of the two-way interactions between variables were significant, with all *p*-values > .605. In contrast, the 3-way interaction between group, type, and story was significant, $F(1, 21) = 6.54$, $p = .018$; $\eta_p^2 = .238$.

**Table 8 Mean proportions (and standard deviations) of complete pronominal forms and nominal forms, by story and by group**

| Story | Referential type | Group | | Both groups |
|---|---|---|---|---|
| | | SLI | TD | |
| April Fools | Pronominals | 0.55 (0.16) | 0.68 (0.12) | 0.61 (0.15) |
| | Nominals | 0.67 (0.18) | 0.82 (0.13) | 0.75 (0.17) |
| Frog on His Own | Pronominals | 0.58 (0.20) | 0.81 (0.12) | 0.70 (0.20) |
| | Nominals | 0.77 (0.11) | 0.84 (0.09) | 0.81 (0.11) |

*Note*. Pronominal forms include both pronouns and ellipsis. One child with SLI never used nominals in *April Fools*, hence for the proportion correct data for *April Fools, n* = 11 for this group.

This significant three-way interaction was followed up with a 2 x 2 mixed ANOVA for each story separately to explore the specific effects of type and of group on referential adequacy. For *April Fools*, the effect of group was significant, with referential adequacy being lower for the SLI group ($M = 1.84$) compared to the TD group ($M = 2.15$), $F(1, 21) = 5.68$, $p = .027$; $\eta_p^2 = .213$. The effect of type was also significant, with referential adequacy being higher for nominals ($M = 2.16$) than for pronominals ($M = 1.84$), $F(1, 21) = 17.3$, $p < .001$; $\eta_p^2 = .452$. The interaction between type and group was nonsignificant, $F(1, 21) = 1.60$, $p = .220$; $\eta_p^2 = .071$. For *Frog on His Own*, the effect of group was significant, with referential adequacy again being

[10] Once again, the marginal means reported correspond to the arcsine transformed values for proportion complete references.

lower for the SLI group ($M$ = 1.96) compared to the TD group ($M$ = 2.32), $F(1, 22) = 8.77$, $p$ = .007; $\eta_p^2 = .285$. The effect of type was also significant, with referential adequacy again being higher for nominals ($M$ = 2.26) than for pronominals ($M$ = 2.02), $F(1, 22) = 12.7$, $p = .002$; $\eta_p^2$ = .367. The interaction between type and group was also significant, $F(1, 22) = 6.38$, $p = .019$; $\eta_p^2$ = .225.

This last interaction between type and group in *Frog on His Own* was followed up with two one-way ANOVAs, one for each type. Results confirmed that for this story, the group difference was nonsignificant for nominals ($M_{SLI} = 2.16$, $M_{TD} = 2.36$), $F(1, 22) = 3.01$, $p = .097$, $d = 0.70$, whereas it was significant for pronominal forms ($M_{SLI} = 1.75$, $M_{TD} = 2.29$), $F(1, 22)$ = 10.6, $p = .004$, $d = 1.33$.

In summary, for *April Fools*, both groups were likewise affected by type, and had lower referential adequacy levels for pronominal forms. Also, overall and for both types, the groups with SLI had lower referential adequacy levels than did their TD peers. This result is somewhat different than what emerged from the prior analysis that considered adequacy by function (see section 3.2.2 above), where the overall effect of group did not quite meet significance. The situation was somewhat different for *Frog on His Own*, as the group with SLI was more affected by type differences, and group differences emerged only for pronominal forms. Hence, depending on the story, the children with SLI had more difficulty clearly referring to story characters either with both types or particularly with pronouns.

## 3.3 The relationship between updating abilities and referential adequacy

The final research question considered whether individual differences in updating abilities were related to referential adequacy in each of the groups, and whether this varied by referential function. The small sample size constrained the statistical analyses that could be considered and reduced statistical power. One solution that was considered to partially overcome these problems was to combine the SLI and TD groups. Visual inspection of the data using scatterplots indicated possibly different relationships between the various memory tasks and referential adequacy for each of the groups. Hence, it was not legitimate to combine the samples. Consequently, the analyses reported here are exploratory at best.

First, one-tailed Spearman ranked correlations were performed between each of the memory tasks and overall referential adequacy for each group separately. Although there were no bivariate outliers in either group,[11] we chose this nonparametric measure given our small sample size and also to avoid the problems caused by high leverage values. See Table 9 for the results of the correlations (also, refer to Table 14 in Appendix D for a summary of the correlations between all memory tasks and referential adequacy, overall and by function, for each group). In the SLI group, none of the updating tasks were significantly correlated with referential adequacy. Only performance on the visual STM task resulted in significant correlations with referential adequacy. In contrast, for the group of TD children, sound monitoring was the only task that correlated significantly with referential adequacy.

**Table 9 Spearman ranked correlations for each memory task with overall referential adequacy, by group**

| Task | Group | | | |
|---|---|---|---|---|
| | SLI | | TD | |
| | $r_s$ | $p$ | $r_s$ | $p$ |
| 1-back | -.151 | .329 | -.035 | .457 |
| Keep track | .372 | .130 | -.415 | .090 |
| Sound monitoring | -.164 | .315 | .536 | .036 |
| Visual STM | .604 | .032 | -.072 | .412 |
| Verbal STM | -.062 | .428 | .422 | .086 |

*Note.* For all but one task, $n$ = 11 for SLI and $n$ = 12 for TD. For the visual span only, $n$ =10 for SLI.

[11] Any data point with a standardised residual value greater than |2.58| was considered an outlier, which corresponds to a 99% confidence interval and is a generally accepted cut-off point for small samples (Tabachnick & Fidell, 2001).

In order to test whether these relationships varied by referential function, these analyses were followed up by two other sets of Spearman ranked correlations that considered maintenance adequacy and switching adequacy separately in terms of their relationships with each of the memory tasks. Following Bamberg (1987), introduction and reintroduction were combined into a switching function. See Table 10 for the results of the correlations. For the children with SLI, none of the memory tasks were significantly correlated with maintenance adequacy, whereas visual STM emerged as the only memory task significantly correlated with switching adequacy. For the TD group, only verbal STM was significantly correlated with maintenance adequacy but no significant correlations surfaced between any of the memory tasks and switching adequacy.

**Table 10 Spearman ranked correlations for each memory task with maintenance adequacy and switching adequacy, by group**

| Task | Maintenance adequacy | | | | Switching adequacy | | | |
|---|---|---|---|---|---|---|---|---|
| | SLI | | TD | | SLI | | TD | |
| | $r_s$ | *p* | $r_s$ | *p* | $r_s$ | *p* | $r_s$ | *p* |
| 1-back | -.452 | .081 | -.041 | .450 | -.091 | .395 | -.048 | .441 |
| Keep track | .405 | .109 | -.202 | .264 | .447 | .084 | -.466 | .063 |
| Sound monitoring | -.425 | .096 | .449 | .072 | -.059 | .431 | .458 | .067 |
| Visual STM | .371 | .146 | -.385 | .108 | .644 | .022 | .109 | .368 |
| Verbal STM | -.293 | .191 | .508 | .046 | .017 | .480 | .463 | .065 |

*Note.* For all but one task, $n$ = 11 for SLI and $n$ = 12 for TD. For the visual span only, $n$ =10 for SLI.

In summary, the two groups showed different patterns of relationships between referential adequacy and WM. On the one hand, for the children with SLI, visual STM was the only memory task that showed any significant correlations with referential adequacy, specifically for referencing overall and for switching, but not for maintenance. The correlation coefficient between referential adequacy and keep track, although nonsignificant, was in the expected direction and of a similar magnitude to the significant correlation found for TD children in the larger sample (Whitely & Colozzo, under review). On the other hand, for the children with TD, there was a relationship between updating and referencing, as sound monitoring was significantly correlated with overall referential adequacy. This result was similar but not identical to that obtained for the larger sample of TD children in Whitely and Colozzo (under review), who found that both sound monitoring and 1-back were significantly correlated with overall referential adequacy and with maintenance. Clearly, a lack of power could be at play here, and therefore it is difficult to make any firm conclusions at this time particularly given the somewhat inconsistent findings when the referential functions were considered separately.

## 3.4 Summary of results

- All updating scores were significantly poorer for the children with SLI compared to the children with TD, with the largest difference found for keep track. Visual STM was the only memory task with nonsignificant differences between groups.
- Both groups used a similar mean total number of references. *Frog on His Own* elicited more referencing by both groups compared to *April Fools*.
- Referential adequacy was significantly different between groups only for *Frog on His Own*, although there was a strong trend in that direction for *April Fools* as well. In both stories, each group was most accurate at referencing for the function of maintenance, and had significantly less success with introduction and reintroduction. Both groups had significantly poorer referential adequacy in *April Fools* compared to *Frog on His Own.* Thus function and story influenced the children in both groups regarding referential adequacy in analogous ways.
- The groups did not differ in distributions of referential types, as they used similar proportions of pronominal forms and nominal forms overall, and for the various referential functions. In terms of function, both groups used a higher proportion of

pronominals for maintenance. Regarding story effects, both groups used a higher proportion of pronominal forms in *April Fools*. Hence, once again function and story affected the children from both groups in comparable ways with regards to the distributions of referential types.

- Using pronominal forms generally resulted in lower referential adequacy than for nominals, although there were subtle differences by story. The children with SLI had more difficulty clearly referencing story characters than did TD peers, but whether this applied to both nominals and pronominal forms or only to pronominals depended on the story.
- For the children with SLI, none of the updating task scores were significantly correlated with referential adequacy; of all the memory tasks, only visual STM scores were significantly correlated with referential adequacy. For the children with TD, only sound monitoring was significantly correlated with referential adequacy. These preliminary analyses, however, are only exploratory. All analyses, and particularly those that considered relationships between updating and referential adequacy depending on referential function, clearly require a larger sample.

# 4 Discussion

This study considered two distinct, yet possibly related, abilities in children with SLI: WM updating and character referencing in narratives. It has the potential to (i) contribute to our understanding of executive functioning in children with SLI; (ii) provide further data to the relatively sparse evidence regarding referential abilities in school-aged children with language difficulties; and (iii) add to the very limited body of research that considers the relationship between WM and language production in any population.

There exists an abundance of research on narrative production by TD children, including studies that focus on cohesion and referencing in particular (e.g., Hickman, 1980; Liles, 1993; Schneider et al., 2006; Wigglesworth, 1990). These skills have been found to improve with age at least into the early school years, although a consensus has yet to be reached with regards to specific ages (Strong & Shaver, 1991; Wigglesworth, 1997; Wong & Johnston, 2004), most likely due to differing task demands (Schneider & Dubé, 1997). There is also a large amount of prior research that has focused on the narrative abilities of children with SLI (e.g., Colozzo et al., 2011; Fey et al., 2004; Liles, 1993), although referencing has received less attention. There is considerable interest regarding the role of WM and of executive functions in language learning and use in both typical and atypical populations, including children with SLI, although the work that specifically considers executive functioning is more recent (Montgomery et al., 2010). While research pertaining to the link between WM and language comprehension has accumulated, investigations of the relationship between WM and language production remain rare (Charest, 2012; Slevc, 2011; Whitely & Colozzo, under review). For children with SLI, there is now accumulated evidence of poorer WM (particularly for verbal tasks) and speed of processing (e.g., Gathercole & Baddeley 1990; Gillam & Hoffman, 2004; Leonard, et al., 2007), but little research has looked at WM updating in either TD children (Carretti et al., 2005;

Palladino et al., 2001; Whitely & Colozzo, under review) or those with SLI (Im-Bolter et al., 2006).

The current study employed three updating tasks that involved visual and auditory presentations and verbal and nonverbal responses. Two STM tasks were also included (one visual and one verbal). Finally, all references to story characters in two elicited narratives were coded for function, adequacy, and type. The current chapter will consider the results relating to each research question in turn. First, group comparisons regarding performance on the memory tasks and the story characteristics will be discussed. Second, the results of group comparisons regarding referential adequacy will be considered, as well as the influences of referential function and of referential type on referential adequacy. Third, possible links between individual differences in WM abilities and in referential adequacy in will be addressed. Finally, the practical implications of these findings and suggestions for future research will be considered.

## 4.1 Memory tasks performance

The tripartite WM model proposed by Baddeley (Baddeley and Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 2003) includes a CE component that supervises the functioning of two subsystems, the phonological loop and the visuospatial sketchpad. Executive functioning is a general control mechanism linked to the CE. Updating is one proposed executive function that monitors incoming information to continuously replace irrelevant information with currently relevant information.

Little research to date has looked at updating abilities in children. Nonetheless, based on the findings of Im-Bolter and colleagues (2006) and on the accumulated evidence regarding WM abilities (Montgomery et al., 2010), the children with SLI were expected to perform significantly below same-age peers on the various updating tasks and the verbal STM task. This was indeed the case. The results generally fit with existing evidence showing that children with SLI have more limited WM abilities (Alloway et al., 2009; Archibald & Gathercole, 2006; Archibald & Joanisse, 2009; Ellis Weismer et al., 1999; Montgomery et al., 2010) than TD peers. Also as expected, the difference between groups was greater for the verbal updating task, keep track.

The children with SLI also performed significantly more poorly than the TD group on the verbal STM task, which has been a consistent finding across studies (Montgomery et al., 2010). The group with SLI also obtained a mean score lower than the TD group for the visual STM task, but this difference was nonsignificant. This fits with the conclusions of Briscoe and Rankin (2009) that visuospatial WM is intact in SLI, although prior evidence is in fact mixed. Some studies have concluded that children with SLI present with both verbal and visual deficits (Hoffman & Gillam, 2004), which suggests domain-general constraints. But many studies have found that the performance on visual tasks tend to be similar between groups, or at least in the average range (Archibald & Gathercole, 2006). Finally, others have proposed that only some of the children with SLI present with visual memory deficits (Alloway et al., 2009). Testing procedures were inconsistent between studies, rendering it difficult to make comparisons.

Performance on STM and WM tasks (Gathercole, 1999) has been found to improve with age in school-aged children with TD. The same has been found in the few studies that have considered WM updating  (Mantyla et al., 2007; Whitely & Colozzo, under review). The poorer memory performance by children with SLI could suggest that their abilities more closely resemble the performance of younger children with TD. Existing research, however, tends to report persistent WM and processing deficits in children with SLI that do not catch up to TD peers with time. Ellis Weismer, Plante, Jones, and Tomblin (2005) investigated verbal WM in 8 children with SLI and 8 children with TD in a longitudinal study that began when the children were in kindergarten. By grade 8, the adolescents with SLI still performed significantly more poorly on verbal WM tasks. In one of the few studies that considered developmental increases in verbal STM tasks in SLI, Conti-Ramsden and Durkin (2007) tested 80 participants with a history of SLI at ages 11 and 14 years. Participants again demonstrated stability in their poor verbal WM over the three-year period. Hence, abilities in the teenagers with SLI had not caught up to the level of TD peers at an age when performance on verbal STM tasks tends to plateau in TD adolescents (i.e., around age 14 or 15). In terms of processing speed in SLI, a review by Montgomery and colleagues (2010) highlighted that the existing literature tends to report slower processing in many children with SLI compared to TD peers through adolescence. Children with SLI do show gains, however, in linguistic and nonlinguistic processing speed between ages 6 and 11 years. No longitudinal studies to date have investigated changes in executive functioning over the long-term in children with SLI. Whether the developmental course for updating will be similar or different to that of other aspects of WM remains an open question.

Research also tends to show that lower overall language abilities persist with increased age in children with SLI (e.g., Leonard et al., 2007; Tomblin, Zhang, Buckwalter, & O'Brien, 2003). For example, Tomblin and colleagues (2003) followed 196 children from kindergarten to grade 4 in a longitudinal study. They found that changes in language status were nonsignificant over these four years. Even 10 years after identifying language impairments, a high percentage of school-aged children did not score more than 1 standard error of measurement higher on formal language tests (Gillam et al., 2008). For this reason, SLI can have negative consequences socially, academically, and vocationally years later (Gillam et al., 2008).

Various explanations for these lasting deficits have been put forth. Leonard and colleagues (2007) suggested that there could be a narrow maturational period for learning specific aspects of language, so that if children with SLI miss this window it is possible that they might never catch up. Tomblin and colleagues (2003) suggested that children with SLI might have limited resources to support their recovery. They hypothesized that the children with SLI who show sufficient improvement have more restricted (versus widespread) deficits to address. Or improvement could simply be an artifact of measurement error caused by regression to the mean by false-positive cases. If certain aspects of memory (i.e., WM and/or executive functioning) do indeed influence language abilities, then deficits or gains in these areas will impact whether language difficulties improve or persist. Archibald and Joanisse (2009) found, however, that only some of the children with SLI had verbal WM impairments, so memory deficits or gains might not affect all children with SLI. There was clearly variability in the current sample of SLI children, as indicated by the large standard deviations, particularly for the visual STM task.

The current study is one of the few to have specifically considered updating abilities in school-aged children. In accord with the results from Whitely & Colozzo (under review), who had adapted the tasks used in the current study to be appropriate with TD 5- to 8-year-olds, the children with SLI were generally able to understand and complete the 1-back, sound monitoring, and keep track tasks. Also in line with previous research, the 2-back task proved very challenging for many of the children, particularly those with SLI.

Caution must be taken in interpreting the results, however, due to the task impurity problem highlighted by van der Sluis and colleagues (2007). Tasks intended to measure executive functioning often involve more than one executive function (e.g., shifting, inhibition), as well as non-executive cognitive abilities (e.g., lexical knowledge). This makes it difficult to be sure that our tasks truly captured the intended executive function (i.e., updating). Further research is necessary to more clearly establish the tasks that will tap a given executive function (Carretti et al., 2005). In addition, studies with larger samples are clearly needed to tease apart the influence of persistent deficits in STM tasks on WM updating in children with SLI. That being said, the tasks used in the current study showed promise in this regard.

## 4.2 Productivity, length, and lexical diversity

Descriptive measures of the narratives varied between the groups as well as by story. Although the children with SLI produced longer stories, their utterances were shorter on average. Similar to younger TD children with less developed discourse skills, this group perhaps adopted simpler strategies as the referential load of storytelling increased the cognitive demands (Emslie & Stevenson, 1980; Wigglesworth, 1997). Alternatively, this could simply reflect their less developed syntactic abilities. Lexical diversity and the number of references were similar between groups. This latter finding that the mean total number of character references produced did not differ between the groups is consistent with prior research (Baltaxe & D'Angiola, 1996; Finestack et al., 2006).

Both groups showed analogous patterns in terms of story differences. *Frog on His Own* elicited longer stories, greater lexical diversity, and more referential tokens than did *April Fools*. The children in both groups were thus comparably influenced and constrained by the stimuli (i.e., the wordless picture books). The story difference in terms of rate of referencing is likely attributable to the fact that *Frog on His Own* involves more characters.

## 4.3 Referential adequacy

### 4.3.1 Overall referential adequacy

Based on prior research (Finestack et al., 2006; Liles, 1985; Norbury & Bishop, 2003; Schneider & Hayward, 2010), children with SLI were not expected to be as successful in clearly referring to characters compared to TD peers. Overall, the SLI group consistently exhibited poorer referential adequacy than did the TD group, but this difference was only significant in *Frog on His Own*. Both groups had lower referential adequacy for *April Fools*; participants had particular difficulty unambiguously referring to the two main characters who were both boys and clearly referring to the groups of people. Thus, TD children performed more similarly to the children with SLI for *April Fools*, the more challenging story.

Difficulties with referencing by the group of children with SLI are likely multi-determined. There is much accumulated evidence that the narrative abilities of children with SLI are not as developed compared to those of peers on several dimensions of form and content, and at both local and global levels (e.g., Colozzo et al., 2011; Fey et al., 2004; Liles, 1993). This could make the task of referencing even more challenging given the cumulative demands of producing a coherent and grammatically accurate narrative in the context of more limited WM resources. In particular, children with SLI may have more difficulty creating and monitoring a discourse model (Johnson-Laird & Garnham, 1980) and thus choosing the appropriate form to meet the needs of the listener (Gundel et al., 1993). Of course, any difficulties with specific forms (e.g., pronouns) would make this even more challenging. Nonetheless, the more similar performance between the groups for the more demanding story suggests that TD children in this age range are also vulnerable to task demands and are continuing to develop their referencing abilities.

### 4.3.2 The influence of referential function on referential adequacy

Adequate referencing is influenced by referential function, as the different functions require (or invite) different forms. The function of introduction generally requires a proper noun or an indefinite article in order to clearly refer to a previously unidentified character for the first time. Maintenance can involve a definite article, a pronoun, or a zero form for subsequent mentions of the same character in adjacent utterances. Reintroduction generally requires a proper noun or a definite article to call attention to a character that was temporarily out of focus.

Results from the referential devices pretest suggest that the group with SLI struggled with pronouns, particular plural forms, which might partially account for the problems referring to groups of people in the narratives. Difficulties with pronouns could affect all referential functions, but may disproportionally impact maintenance. As far as the indefinite/definite distinction, this continues to challenge TD and SLI children alike in this age-range although they are clearly able to use both forms in other contexts. Both the indefinite and definite determiners were correctly elicited from all children with SLI on the referential devices pretest, indicating that errors with these forms in the narratives were not attributable to them not having been acquired yet.

The third research question considered whether the difficulties that children with SLI and TD have in referential adequacy varied systematically and in similar ways depending on the referential function. Although little research involving children with SLI has compared the referential functions, the children in both groups were expected to be similarly influenced by function, and specifically to have the highest levels of referential adequacy for maintenance, and the lowest for reintroduction. Results in the current study supported these predictions. There were, in fact, more similarities than differences between groups, suggesting that referential function and story features influenced all children in analogous ways. The three referential functions did not entail equal levels of difficulty. Referential adequacy was significantly higher for maintenance compared to either of the switching functions in both stories and for both groups. However, reintroduction proved more challenging than introduction only for one of the stories, *April Fools*, and this was again the case for both groups (i.e. there was no interaction between function and group). This story includes a dyad (the two boys) as well as groups of people (i.e., the townspeople) that are often, but not always, sharing activities. These features of the story seem to have made it more difficult for the children to unambiguously refer to the characters, particularly when switching. This led to more similar performance between the groups for introduction and for reintroduction, which likely accounts for the nonsignificant difference between the groups for overall adequacy in *April Fools*.

### 4.3.3 The influence of referential type on referential adequacy

Differences in referential adequacy between the groups could have been driven by differences in types, although the limited existing research suggested that both groups of children would present with similar patterns of use for the various referential types (i.e., nominals vs. pronominals/ellipsis). Alternatively, these differences in adequacy could be attributable to relatively more difficulties with some forms compared to others. In particular, the referential devices pretest indicated that some children with SLI continued to struggle with pronouns.

The fourth research question addressed whether referential type distributions were similar between the SLI and TD groups, and how type interacted with referential adequacy. Results indicated that the groups used similar proportions of pronominal forms overall and for each of the functions. The groups were once again influenced by function in the same way, with pronouns being used more frequently for maintenance compared to either of the switching functions. This result suggests that the children in both groups were sensitive to listeners' needs. The groups were also once again comparably influenced by story, which proved to be the more influential factor: the proportion of pronominals was much higher in *April Fools* ($M_{\text{SLI}}$ = .73; $M_{\text{TD}}$ = .82) compared to *Frog on His Own* ($M_{\text{SLI}}$ = .44; $M_{\text{TD}}$ = .46), and this was true across functions.

In terms of the influence of type on referential adequacy, the story with the lower overall referential adequacy levels, *April Fools*, caused problems for children in both groups mostly due to the overuse of pronouns, which had a particularly detrimental effect on adequacy for switching (groups combined, $M_{\text{AF}}$ = .43, $M_{\text{FOHO}}$ = .65). Although the two groups were likewise affected by story, and the group with SLI consistently showed lower adequacy levels for both types, the results were not identical for both stories. Whereas for *April Fools* the children with SLI had significantly poorer referential adequacy than the children with TD for nomimal as well as pronominal forms, in *Frog on His Own*, the group difference was significant only for pronominals. Hence whether the children with SLI had more difficulty clearly referring to story characters either with both types or particularly with pronouns depended on the story, with the more demanding story resulting in significant differences regardless of type.

## 4.4 The relationship between updating and referential adequacy

Storytelling is a complex task that requires a speaker to construct a plot, consider the context, track what has been said, and monitor the listener's knowledge. This draws upon linguistic, social, pragmatic, and cognitive resources. Existing literature proposes that there should be a strong relationship between memory development and the development of adequate use of referential devices (Karmiloff-Smith, 1985). Referencing in narratives requires the speaker to constantly monitor and revise relevant information, which would place particular demands on updating abilities. As mentioned above, the various referential functions have proven in past research to be more or less difficult for children to clearly refer to characters (Whitely & Colozzo, under review). The functions are also likely to tax a storyteller's updating resources to a different degree. Introduction should be relatively less demanding, particularly for character introductions that take place early in the story when there are fewer characters to manage and a less elaborate discourse model to update. Speakers must nonetheless evaluate the listener's knowledge and attention, and then choose an appropriate linguistic device to clearly mention a new character for the first time. As more characters have entered the story, the narrator must keep track of an increasing number of them for unambiguous referencing. This could make subsequent mentions more challenging. Nonetheless, one would expect maintenance to make the least demands on updating, since the same character remains in focus as the speaker refers to it in contiguous utterances. Finally, character reintroduction should be the most challenging function due to the greater time and distance between mentions (Daneman &

Carpenter, 1980). Intervening information and characters will thus compete with those that are being reintroduced and that have been temporarily outside of the focus of attention.

The fifth research question asked whether individual differences in updating abilities in WM were related to referential adequacy in children with SLI and TD, and whether this interacted with referential function. Based on the specific task demands, the findings of Whitely and Colozzo (under review) for TD children, and other related research that looked at updating and language-based tasks (Carretti et al., 2005; Palladino et al., 2001; van der Sluis et al., 2007), significant relationships were expected between updating and referencing in both groups. The analyses were limited, however, by the small sample size. As a result, this research question could not be addressed in depth. The results must be treated with caution, and serve only as a tentative guide for future research.

Contrary to expectations, for the children with SLI there were no significant correlations between scores for any of the updating tasks and referential adequacy, although the correlation coefficient for keep track was in the expected direction and larger than the other coefficients. Of the two STM tasks, again surprisingly, only scores for the visual STM task were significantly correlated with referential adequacy (overall and for the function of switching). None of the memory tasks were significantly correlated with maintenance adequacy. For the children with TD, only one updating task (sound monitoring) was significantly correlated with overall referential adequacy. No updating tasks were significantly correlated with adequacy when analyzed by function.

There are many possible explanations for these unexpected results. First, various executive and non-executive cognitive factors influence performance on any task, and character referencing in narratives is no exception. This fact has led to a high proportion of variance in the performance on complex cognitive tasks that remains unaccounted for in many studies (Miyake et al., 2000). In the current case, the performance of the children with SLI in particular in terms of referential adequacy may have been influenced by non-executive resources and linguistic variables to an extent that made it difficult to isolate the effect of WM updating. This is even more likely given the small sample. Second, it is well established that the deficits in SLI are heterogeneous, so the relationship could be stronger for some but not all children with SLI. In fact, not all children with SLI have exhibited deficits in WM (e.g., Archibald & Joanisse, 2009). For example, although Im-Bolter and colleagues (2006) found that processing limitations in children with SLI were mediated by poorer updating abilities, the authors pointed out that performance varied greatly between tasks and individuals. In the current study, the standard deviations were also often quite large, not only for the WM tasks but also for referential adequacy. This again speaks for the need to increase sample size. Third, of the few studies that have found a significant role for updating in language abilities, most have been based on TD children, many of whom were older than the children in the current study (e.g., ages 8 to 11 years in Carretti et al., 2005; ages 9 to 12 years in van der Sluis et al., 2007). The extent to which those results generalize to younger children has yet to be determined. Fourth, various statistical issues could have clouded the analysis. The small sample size resulted in low power of the statistical tests and made it impossible to use multiple regression analyses, which would have allowed to better tease apart the effects of storage capacity, language ability, and updating. In the sample of TD children in the study by Whitely and Colozzo (under review), the younger children showed a stronger influence of the non-executive memory tasks compared to the older children. Children with SLI and young (i.e., 5-year-old) TD children alike may rely more on their storage capacity to keep track of characters during their narrative production, whereas for older children, WM updating may play a larger role. At this time, these ideas are little more than interesting conjecture. Finally, the finding that the visual STM task correlated significantly with overall referential adequacy for the children with SLI is intriguing and requires further investigation.

This could have to do with the larger number of trials for the visual STM tasks that better reflected individual differences than did the verbal STM task.

Whitely and Colozzo (under review) also considered the influence of specific functions on the relationship between WM updating on referential adequacy, and found that the strongest relationship emerged for maintenance. This result led the authors to conclude that executive processes are involved when tasks are sufficiently but not excessively challenging. How the relationship between updating and referencing will manifest itself in children with SLI remains an open question. It is possible that a larger sample of children will show equally strong relationships with updating across all referential functions. Together, limitations in language knowledge and WM resources could result in all of the functions being more similarly demanding in the SLI group. It is also feasible that the effect of updating beyond that of non-executive resources will emerge only in older children.

## 4.5 Practical implications

The above findings highlight important things to keep in mind when assessing, diagnosing, and treating children with SLI. The potential impact of WM and processing capacity should be considered at all stages of intervention (Montgomery et al., 2010). For assessment, the root of the issue and not just the outcomes of skills that are evident on the surface should be identified (Alloway et al., 2009). If referencing is a particular problem, which has been found in research on SLI, it is important to clarify whether the issue can be traced to the acquisition of referential forms, or to the use of these forms in demanding tasks such as narrative. Language performance on standardized tests or more naturalistic classroom activities could be negatively affected by memory limitations as well. In addition to linguistic abilities, cognitive abilities and participation should be dynamically assessed in natural contexts to capture a child's strengths and weaknesses (Gillam & Hoffman, 2004).

Intervention should be tailored to a child's individual profile (Alloway et al., 2009). Children with SLI might require specific training of language form and/or content at both local and global levels (Colozzo et al., 2011). In addition to treating language deficits, targeting cognitive skills to improve language performance should be considered. This can involve the provision of specific training, compensatory strategies, or external memory cues or aids (Carretti et al., 2005; St Clair-Thompson & Gathercole, 2006). Teachers or therapists can also provide support by decreasing task demands, such as the size of the memory load and the required speed of processing, to prevent overloading the child's system. Just as longer narratives with greater lexical diversity and referential adequacy were elicited for *Frog on His Own*, which is considered a less demanding story compared to *April Fools*, decreasing the cumulative cognitive demands of a task for children with more limited memory resources might support language production and offer a better starting point to build on for new skills. Both linguistic and memory abilities heavily affect a child's academic skills in the long-term, so thorough assessment and intervention is critical to a child's potential success (Montgomery et al., 2010).

## 4.6 Future directions

This study considered two factors in children with SLI: WM updating abilities and character referencing in narratives. It is the first to explore the relationship between updating and referencing in the narratives produced by children with SLI, and one of the few to consider links between WM and language production in any group. However, due to the small sample size, more questions than answers have resulted regarding the possible links between WM updating and referencing in these children. This study is nonetheless promising in that the 5- to 8-year-old children with SLI were generally able to understand and complete the 1-back, sound monitoring, and keep track tasks. Hence, these tasks show promise for future work that considers WM updating in this population.

There are many interesting directions researchers can go from this point. The most obvious launch point would be to pursue this research by increasing the sample size. This would increase statistical power and make it possible to consider other statistical procedures, including partial correlations and multiple regression. Whitely & Colozzo (under review) were the first to explore the relationship between WM updating and a language production task in TD children, as the role of WM had previously been studied mostly in relation to language comprehension. Many more studies would be required to develop a more complete understanding of the relationship between executive functioning and language production.

# 5 Conclusion

This study has contributed to our understanding of executive functioning in school-aged children with SLI aged 5 to 8 years as well as provided greater insight into their referencing abilities. The children with SLI performed more poorly than did same-age TD peers on all three updating tasks, with the most verbal task showing the largest group difference. In terms of referencing, there were many more similarities than differences between the groups. Specifically, children with and without language impairments responded in comparable ways to both story differences and to the demands of the various referential functions in terms of the frequency of character references and the referential types selected. They also showed analogous patterns in terms of referential adequacy, with greater success with maintenance compared to either of the switching functions and lower levels of referential adequacy for *April Fools* than for *Frog on His Own*. The children with SLI did, nonetheless, generally have lower referential adequacy levels than their TD peers, but whether this applied to both nominal and pronominal forms depended on the story.

Regarding the relationship between WM and referencing, results indicated that only the visual STM task (but none of the updating tasks) was significantly related to referential adequacy for this small sample of children with SLI. The specific mechanisms involved in the ability to adequately refer to characters in storytelling still remain unclear. It is possible that updating does play a role, but this relationship could be mediated or overshadowed by the non-executive WM resources or linguistic variables for children with SLI in this developmental window. Due to the small number of participants in this study, the results regarding the relationship between updating and referential adequacy are tentative and require replication with a larger sample.

## Bibliography

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health, 86,* 726-728.

Alloway, T. P., Rajendran, G., & Archibald, L. M. D. (2009). Working memory in children with developmental disorders. *Journal of Learning Disabilities, 42,* 372-382.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders – Text revision* (4th ed.). Washington, DC: Author.

American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening* [Guidelines]. Available from www.asha.org/policy.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language and Communication Disorders, 41,* 675-693.

Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52*, 899-914.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). San Diego, CA: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 28-61). Cambridge, UK: Cambridge University Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Baltaxe, C. A. M. & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders, 22,* 1-21.

Baltaxe, C. A. M., & D'Angiola, N. (1996). Referencing skills in children with autism and specific language impairment. *European Journal of Disorders of Communication, 31*, 245.

Bamberg, M. (1987). *The acquisition of narratives: Learning to use language.* Berlin: Mouton de Gruyter.

Beliavsky, N. (2003). The sequential acquisition of pronominal reference in narrative discourse. *WORD, 54,* 167-189.

Bennett-Kastor, T. (1983). Noun phrases and coherence in child narratives. *Journal of Child Language, 10,* 135-149.

Botting, N., Faragher, B., Simkin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: What differentiates good and poor outcome? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 42*, 1013-1020.

Briscoe, M., & Rankin, P. M. (2009). Exploration of a 'double-jeopardy' hypothesis within working memory profiles for children with specific language impairment. *International Journal of Language & Communication Disorders, 44,* 236-250.

Carretti, B., Cornoldi, C., De Beni, R., & Romano, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology, 91*, 45-66.

Charest, M. J., & Johnston, J. R. (2011). Processing load in children's language production: A clinically oriented review of research. *Canadian Journal of Speech-Language Pathology and Audiology, 35,* 18-31.

Charest, M. J. (2012). *Lexical activation effects on children's sentence production*. (Unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada.

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 54,* 1609-1627.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language and literacy: Developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry, 48,* 147-156.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450-466.

Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1249-1260.

Ellis Weismer, S., Plante, E., Jones, M., & Tomblin, J. B. (2005). A functional magnetic resonance imaging investigation of verbal working memory in adolescents with specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 405-425.

Emslie, H. C., & Stevenson, R. J. (1980). Preschool children's use of the articles in definite and indefinite referring expressions. *Journal of Child Language, 8*, 313-328.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1301-1318.

Finestack, L. H., Fey, M. E., & Catts, H. W. (2006). Pronominal reference skills of second and fourth grade children with language impairment. *Journal of Communication Disorders, 39*, 232-248.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201-225.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336–360.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39*, 3-27.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3,* 410-419.

Gillam, R. B., & Hoffman, L. M. (2004). Information processing in children with specific language impairment. In L. Verhoeven & H. van Balkom (Eds.), *Classification of developmental language disorders: Theoretical issues and clinical implications.* (pp. 137-157). Mahwah, NJ: Erlbaum.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., … Friel-Patti, S. (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 51,* 97-119.

Grice, P. (1975). Logic in conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Vol. 3: Speech Acts.* (pp. 41-58). New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive statuses and the form of referring expressions in discourse. *Language, 69,* 274-307.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, speech, and hearing services in schools, 41,* 84-95.

Hickman, M. (1980). Creating referents in discourse: A developmental analysis of linguistic cohesion. In J. Kreiman & A. Ojeda (Eds.), *Papers from the parasession on pronoun and anaphor.* Chicago, IL: Chicago Linguistic Society.

Hoffman, L. M., & Gillam, R. B. (2004). Verbal and spatial information processing constraints in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 114-125.

Im-Bolter, N. Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development, 77*, 1822-1841.

Isard, S. (1974). 'Changing the context'. In E.L. Keenan (Ed.), *Formal Semantics of Natural Language,* Cambridge University Press, Cambridge.

Johnson-Laird, P.N., & Garnham, A. (1980). Descriptions and discourse models. *Linguistics and Philosophy, 3,* 371-393.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach* (2nd ed.). New York, NY: Routledge.

Karmiloff-Smith, A. (1980). Psychological processes underlying pronominalization and non-proniminalization in children`s connected discourse. In J. Kreiman and A. E. Ojeda (Eds), *Papers from the parasession on pronouns and anaphora.* Chicago, IL: Chicago Linguistics Society.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes, 1,* 61-85.

Kaufman, A. S., & Kaufman, N. L. (2004). *KBIT-2: Kaufman brief intelligence test, 2nd Ed*. San Antonio, TX: Pearson.

Krahn, F. (1974). *April fools.* New York, NY: Dutton.

Lehto, J. H. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology, 49A,* 29-50.

Leonard, L. B., Ellis Weismer, S., Miller, C. A., Francies, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research, 50,* 408-428.

Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*, 123-133.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research, 36*, 868-882.

Loban, W. (1976). *Language development: K through Gr. 12. Research Report No. 18*. Urbana, IL: National Council of Teachers of English.

Logie, R. H., & Pearson, D. G. (1997). The inner-eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology, 9*, 241-257.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York, NY: Psychology Press.

McNiven, C. L. (2010). *Who's who: Memory updating and character reference in children's narratives.* (Unpublished master's thesis). University of British Columbia, Vancouver, Canada.

Mantyla, T., Carelli, M. G., & Forman, H. (2007). Time monitoring and executive functioning in children and adults. *Journal of Experimental Child Psychology, 96*, 1-19.

Maratsos, M. P. (1974). Preschool children's use of definite and indefinite articles. *Child Development, 45,* 446-455.

Mayer, M. (1973). *Frog on his own.* New York, NY: Dial Press.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*, 30-43.

Miller, J., & Inglesias, A. (2006). Systematic Analysis of Language Transcripts (SALT), English and Spanish (Version 9) [Computer Software], Language Analysis Lab, University of Wisconsin-Madison.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

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.

Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development-Primary: Fourth Edition (TOLD-P:4)*. Austin, TX: Pearson.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*, 287-313.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46–59.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29,* 344-354.

Roth, F. P., Spekman, N. J., & Fye, E.C. (1995). Reference cohesion in the oral narratives of students with learning disabilities and normally achieving students. *Learning Disability Quarterly, 18,* 25-40.

Schneider, P., & Dubé, R. (1997). Effect of pictorial versus oral story presentation on children's use of referring expressions in retell. *First Language, 17,* 283-302.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 1.1) [Computer software]. Pittsburgh, PA: Psychological Software Tools.

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology, 30,* 224-238.

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41,* 459-473.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, fourth edition (CELF-4).* Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1503-1514.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 54,* 745-759.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*, 407-418.

Strong, C. J., & Shaver, J. P. (1991). Stability of cohesion in the spoken narratives of language-impaired and normally developing school-aged children. *Journal of Speech and Hearing Research, 34,* 95-111.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Boston: Allyn and Bacon.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39,* 1284-1294.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. Journal of Speech, Language, and Hearing Research, 46, 1283-1296.

Van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35,* 427-449.

Warden, D. A. (1976). The influence of context on children's use of indentifying expressions and references. *British Journal of Psychology, 67,* 102-112.

Whitely, C., & Colozzo, P. (under review). Who's who? Memory updating and character reference in children's narratives. *Journal of Speech, Language, and Hearing Research*.

Wigglesworth, G. (1990). Children's narrative acquisition: a study of some aspects of reference and anaphora. *First Language, 10*, 105-125.

Wigglesworth, G. (1997). Children's individual approaches to the organization of narrative. *Journal of Child Language, 24*, 279-309.

Winer, B. J. (1991). *Statistical principles in experimental design* (3rd ed.). New York, NY: McGraw-Hill.

Wong, M. Y. A., & Johnston, J. R. (2004). The development of discourse referencing in Cantonese-speaking children. *Journal of Child Language, 31*, 633-660.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors, 5*, 7-17.

# Appendices

## Appendix A: Standardized Language and Cognitive Test Scores

**Table 11 Standard scores for each child with SLI on the TOLD-P:4 and the KBIT-2**

| Subject | TOLD-P:4 Subtests | | | | | | TOLD-P:4 Spoken Language Composite | KBIT-2 Standard Scores |
|---|---|---|---|---|---|---|---|---|
| | PV | RV | OV | SU | SI | MC | | |
| 1501 | 6 | 7 | 1 | 8 | 6 | 3 | 67 | 124 |
| 1502 | 11 | 5 | 4 | 8 | 5 | 8 | 78 | 83 |
| 1503 | 6 | 2 | 1 | 10 | 5 | 5 | 65 | 89 |
| 1504 | 4 | 2 | 1 | 4 | 1 | 3 | 50 | 75 |
| 1602 | 10 | 7 | 4 | 8 | 7 | 7 | 81 | 85 |
| 1603 | 13 | 11 | 4 | 13 | 13 | 3 | 96 | 126 |
| 1604 | 9 | 3 | 6 | 7 | 1 | 4 | 66 | 105 |
| 1605 | 14 | 10 | 7 | 8 | 5 | 9 | 92 | 102 |
| 1702 | 13 | 7 | 7 | 8 | 6 | 8 | 86 | 97 |
| 1703 | 12 | 4 | 6 | 10 | 7 | 7 | 84 | 105 |
| 1704 | 10 | 8 | 6 | 9 | 5 | 9 | 85 | 91 |
| 1705 | 8 | 8 | 7 | 13 | 7 | 8 | 89 | 115 |
| Mean | 9.67 | 6.17 | 4.50 | 8.83 | 5.67 | 6.17 | 78.25 | 99.75 |
| SD | 3.17 | 2.98 | 2.39 | 2.48 | 3.08 | 2.41 | 13.54 | 16.18 |

*Note*. PV = Picture Vocabulary; RV = Relational Vocabulary; OV = Oral Vocabulary; SU = Syntactic Understanding; SI = Sentence Imitation; MC = Morphological Completion.

## Appendix B: Referential Devices Pretest Performance

**Table 12 Performance by the children with SLI for each target in the referential devices pretest**

| Subject | my | z | their | his | hers | it | them | him | her | he | she | I | me | they | you | a | the | who/ that |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1501 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1502 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1503 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| 1504 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 1602 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1604 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1605 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 1702 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| 1703 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| 1704 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 1705 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |

*Note*: data is missing for one child. 1 = child correctly produced the referential form; 0 = the child did not produce the referential form.

## Appendix C: Raw Data

**Table 13: Age (months) and scores for each participant on the memory tasks and for referential adequacy, by story**

| Subject ID | Age | Keep Track | 1-back | 2-back | Sound monitoring | Visual span | Digit span | Reference AF | Reference FOHO |
|---|---|---|---|---|---|---|---|---|---|
| SLI | | | | | | | | | |
| 1501 | 67 | 4 | .81 | .57 | .70 | 10 | 5 | .40 | .54 |
| 1502 | 66 | 10 | .76 | .63 | .67 | 7 | 3 | .48 | .69 |
| 1503 | 67 | 6 | .87 | .56 | .59 | 4 | 4 | .67 | .73 |
| 1504 | 77 | 7 | .59 | .59 | .67 | 0 | 1 | .28 | .43 |
| 1602 | 78 | 10 | .65 | .74 | .81 | 27 | 6 | .81 | .84 |
| 1603 | 84 | 7 | .94 | .67 | .96 | 8 | 8 | .29 | .60 |
| 1604 | 89 | 8 | .91 | .61 | .67 | 12 | 3 | .61 | .76 |
| 1605 | 86 | 9 | .87 | .72 | .81 | 11 | 5 | .52 | .61 |
| 1702 | 89 | 12 | .91 | .69 | .93 | 15 | 5 | .78 | .64 |
| 1703 | 89 | 11 | .85 | .65 | .74 | 21 | 5 | .80 | .91 |
| 1704 | 95 | 7 | .93 | .67 | .78 | | 5 | .65 | .78 |
| 1705 | 97 | 12 | .98 | .69 | 1.00 | 13 | 9 | .60 | .72 |
| TD | | | | | | | | | |
| 55 | 70 | 12 | .91 | .61 | .89 | 9 | 5 | .56 | .83 |
| 514 | 67 | 13 | .91 | .78 | .93 | 10 | 5 | .71 | .60 |
| 58 | 68 | 13 | .93 | .74 | 1.00 | 21 | 7 | .78 | .80 |
| 523 | 77 | 14 | .94 | .70 | .67 | 23 | 8 | .49 | .73 |
| 610 | 81 | 8 | .87 | .70 | .93 | 17 | 5 | .67 | .77 |
| 613 | 84 | 8 | .91 | .67 | .93 | 6 | 7 | .85 | .87 |
| 67 | 84 | 16 | .96 | .78 | .93 | 15 | 8 | .65 | .88 |
| 624 | 88 | 6 | .89 | .63 | 1.00 | 13 | 9 | .86 | .90 |
| 71 | 95 | 21 | .93 | .67 | .96 | 14 | 7 | .61 | .73 |
| 78 | 91 | 17 | .98 | .81 | 1.00 | 24 | 8 | .68 | .91 |
| 712 | 98 | 17 | .96 | .74 | .93 | 20 | 6 | .73 | .90 |
| 719 | 98 | 12 | .94 | .69 | .96 | 17 | 8 | .81 | .89 |

*Note*. Keep track, total number of items correct. 1-back, 2-back, and sound monitoring, proportion correct *p(c)** = (hits + correct rejections)/total. Visual span and digit span, total number of trials correct. Reference, proportion correct for *April Fools* (AF) and *Frog on His Own* (FOHO).

## Appendix D: Summary of Correlations

**Table 14: Spearman ranked correlations between referential adequacy (overall, maintenance, and switching), updating, and STM span measures for each group**

| Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| SLI | | | | | | | | |
| 1. Overall adequacy | - | .806** | .995** | -.151 | .372 | -.164 | .604* | -.062 |
| 2. Maintenance adequacy | | - | .782** | -.452 | .405 | -.425 | .371 | -.293 |
| 3. Switching adequacy | | | - | -.901 | .447 | -.059 | .644* | .017 |
| 4. 1-back | | | | - | .078 | .509 | -.079 | .378 |
| 5. Keep track | | | | | - | .494 | .659* | .257 |
| 6. Sound monitoring | | | | | | - | .406 | .865** |
| 7. Visual span | | | | | | | - | .408 |
| 8. Verbal span | | | | | | | | - |
| TD | | | | | | | | |
| 1. Overall adequacy | - | .743** | .923** | -.035 | -.415 | .536* | -.072 | .422 |
| 2. Maintenance adequacy | | - | .522* | -.041 | -.202 | .449 | -.385 | .508* |
| 3. Switching adequacy | | | - | -.048 | -.466 | .458 | .109 | .463 |
| 4. 1-back | | | | - | .807** | .160 | .657* | .359 |
| 5. Keep track | | | | | - | .117 | .480 | .029 |
| 6. Sound monitoring | | | | | | - | .187 | .469 |
| 7. Visual span | | | | | | | - | .324 |
| 8. Verbal span | | | | | | | | - |

*Note.* For all but one variable, $n$ = 11 for SLI and $n$ = 12 for TD. For the visual span only, $n$ =10 for SLI. $^{**}p \leq .01$, $^{*}p \leq .05$ (1-tailed).

## Appendix E: Detailed Scoring Decisions for Referential Adequacy by Referential Function

### Referential function

Introduction: first mention of a story character

- Complete: proper noun or indefinite article + noun
- Incomplete: pronoun or definite article + noun. Exceptions: pronouns specified cataphorically, possessive pronouns (e.g., *his* dog).

Maintenance: successive reference to the same character

- Complete: pronoun, proper noun, definite article + noun, or zero form (ellipsis)
- Ambiguous: pronoun with more than one possible referent or indefinite article + noun

Reintroduction:

- Complete: proper noun or definite article + noun
- Incomplete: pronoun (unless forming a contiguous string with the immediately preceding utterance)
- Ambiguous: indefinite article + noun, a pronoun with more than one possible referent, or failure to specify a character's identity at each scene change.

### Additional considerations

- The first mention in a new scene needed to be specified with a noun phrase, unless it was specified in the immediately preceding utterance.
- Any reference error (e.g., in character name, switch in number, gender, or person) was coded as ambiguous.
- Inappropriate ellipsis was coded as ambiguous.

- An unspecific pronoun (e.g., everybody, people, someone) that was not disambiguated in the scene (specified cataphorically), or not used in a truly generic way, was coded as ambiguous.
- If there was a change from multiple characters being referred to as a group to singling out one character, or vice versa, it was considered reintroduction (Bamberg, 1987).
- In *Frog on His Own*, the speaker needed to reintroduce the frog's owner and his dog in the final scene with additional information (more than a definite article + noun, which would be ambiguous), because they had not been mentioned since the first scene.
- If the child erroneously interpreted a character's identity, it was coded according to the child's interpretation.
- The pictures in the story and the video of the child's physical gestures (exophoric referencing) were noted and used to supplement coding decisions.

## Appendix F: Story Details

**Scenes**

*April Fools*

1. The boys build a monster.
2. The boys scare a sleeping woman (victim 1).
3. The townspeople see a monster on the rooftop of a house and the boys scare a man in his house (victim 2).
4. The boys take the monster across the river and get lost in the forest. The townspeople find them.

*Frog on His Own*

1. The boy and his pets, including the frog, arrive in the park. The frog escapes.
2. The frog catches a bee.
3. The frog disrupts a couple's picnic.
4. The frog breaks a boy's boat.
5. The frog jumps in a baby's carriage and gets chased by a cat. The frog is saved by his owner and the dog, and they leave the park.

**Characters**

*April Fools*

- Boy 1
- Boy 2
- Monster
- Townspeople
- Victim 1 (sleeping old lady)
- Victim 2 (man in doorway)

*Frog on His Own*

- Boy 1 (with pets)
- Frog
- Dog
- Turtle
- Fly
- Bee
- Man (picnic)
- Woman (picnic)
- Boy 2 (boat)
- Mom 1 (boat)
- Baby
- Mom 2 (with baby)
- Cat

## Appendix G: Detailed Task Instructions

**TOLD – 4 Ed.**

"We are going to start by looking at a book with a lot of pictures where I need you to point with your finger, and we are also going to talk about some words." Administer the Core Language Subtests following all directions as indicated by the TOLD-P:4 scoring sheet.

**Kbit-2**

"We are going to look at a book with some more pictures. All you have to do is point to a picture." Administer only the Matrices Subtest following all directions as indicated by the Kbit-2 scoring sheet.

**Visual Span**

"We are going to start with an activity on the computer. In this game you will see squares. Some of the squares are going to be red. Then all of the squares will disappear, and when they come back one of the red squares will be gone. Your job is to point to the square that was red before. There are going to be more and more squares as we go. If you're not sure, it is ok to guess. Just do your best. Ready?" Complete the practice. If the child makes a mistake, model the correct response and have the child do the practice again. Discontinue when the child scores 0 or 1 out of 3 at any level.

**Hearing Screening**

"We are going to play a listening game now. I'm going to put these headphones over your ears, and you need to listen for quiet beeps. Every time you think you hear a beep, clap your hands. The beeps will get very quiet, so make sure to clap if you think you heard something." Practice with the child at 50 dB, then use 30 dB to orient the child to each frequency level.

**Digit Span**

"I'm going to say some numbers. Listen carefully, and when I'm finished, you say them in the same order as I said them. For example, if I say 1, 2, you say 1, 2." Pause for 1 second between each number. Discontinue when the child responds incorrectly on both items on a single level.

**Reference Pretest**

"We are going to look at some pictures. I am going to start the sentence and I need you to finish it. Try your best!" Open the blue binder and place it in front of the child. Say each phrase to the child and record their answer below. Only administer the first probe. If the child provides an incorrect response to the first probe, say, "That's right we could say "___" (hers), but for this picture we need to say "___" (Holly's). Let's try this one again." Administer the second probe, and if the child responds incorrectly a second time, say, "Great work, let's keep going!"

**1-back**

"This is a computer game. You will see three dots on the screen. These dots can line up in different ways. Here is what they look like." Show the 10 possible dot configurations on the computer.

"In this game your job is to watch for a picture of the dots that is the same as the picture that came just before it. Let's try some for practice. Look (present card 1), remember this (flip over card 1, present card 2). Here's the next one. Is it the same as the one before? Right, it's not the same. Now remember this one (point to card 2 then flip it over, and present card 3). Is this one the same as the one before? Now remember this one (point to card 3, flip it over, and present card 4). Is this one the same as the one before?" Continue flipping over each card and asking the child yes or no.

“Now let’s do it on the computer. I’m going to do some first, and you watch. I am going to press this button (point to the space bar) if the picture of the dots is the same as the picture just before it. So look, remember this. Is this one the same as the one before?” Repeat for each array and nod or shake your head accordingly.

“Ok, now it’s your turn to do it on the computer. I want you to press this button (point to the space bar) if the one you see is the same as the one that came before it. Just try your best. If you are not sure it’s ok to guess.” Present the practice set on the computer. If the child does not get them all correct, repeat the practice again with them: “That was close, let’s do it one last time together.”

“Good job!  Now you’re going to do this game 3 times by yourself.” Complete the three trials. If a break is needed between trials, repeat the instructions” “Your job is to press this button (point to space bar) if the picture of dots on the screen is the same as the picture before it.”

**Sound Monitoring**

Put headphones on both the experimenter and the child. “We are going to play a game with sounds. These are the three sounds.” Play the three sounds for the child without naming them. “You are going to hear these sounds a bunch of times. Wait until you hear the same sound three times, and then press this button (point to the space bar). I will do the first one to show you. Ready?” The experimenter completes the demonstration as the child listens. The child then completes the practice. If the child makes any mistakes, the experimenter repeats the practice with the child” “That was close! Let’s do it one last time together. Good job. Now you will do it by yourself. Wait until you hear the same sound three times, then press the button.”

**Keep Track**

“We are going to play a computer game. In this game there are going to be a lot of different things that belong in six groups.” Show each icon as you name the category. “Now I am going to show you all of the things that go in these groups. You tell me what it is and then click on the group that the picture belongs to. Ready?” Ask, “What’s this?” as necessary to elicit a response. If the child makes an error, stop and redo that object.

“Now I am going to show you how to play this game. My two groups are animals and clothing. We are going to see some of the pictures. At the end I have to say the last animal that I saw and the last piece of clothing that I saw, so those are the ones I have to remember. I am going to do this one, you watch, and then it will be your turn.”

“Now it is your turn. Your two groups are body parts and toys. At the end you have to tell me the last body part that you saw and the last toy that you saw. Ready?” If the child makes any mistakes, say, “That was so close, let’s try it again” and redo the practice once more with the child. “Great job! Now we are going to do another one. This time your groups are animals and clothing. Ready?” Go through the first six blocks. Provide the child with the categories before each block begins.

At block 7 say, “Now it is going to get a little bit tricky. You have to remember three groups. This time your groups are transportation, animals, and toys. At the end, try to tell me the last transportation picture that you saw, the last animal that you saw, and the last toy that you saw. Ready?”

**April Fools**

"I'm going to show you a book that has only pictures, and no words. I want you to look through the pictures and think of a story to tell that goes with the pictures. When you are ready you can start telling your story. We are recording your story for my friend, Cristy. She will listen to it later, but she won't have the book. Tell me when you are ready." The child looks at each picture in the book. "Ok now tell the best story you can and I'll record it for my friend Cristy."

**Frog on His Own**

"Now we are going to tell another story. This time we are going to use this book. It is another book that has no words, and only pictures. I want you to look through the pictures and think of a story to tell that goes with the pictures. When you are ready you can start telling your story. Just like before, we are recording this story for my friend, Cristy. She will listen to it later, but she won't have the book. Tell me when you are ready." The child looks at each picture in the book. "Ok now tell the best story you can and I'll record it for my friend Cristy."

**2-back**

"This is a dot game, kind of like the one we played before, but the rules are going to be a little bit different. In this game, your job is to watch for a picture of the dots that is the same as the picture that came two pictures before it. Let's try some for practice. Look (present card 1), remember this (flip over card 1, present card 2). We have to remember this one too (point to card 2) and we're still remembering this one (point to card 1). Now look at this one (flip over card 3), is it the same as the one you saw two pictures ago (point to card 1)? We are remembering this one (card 2), want to peek? Now we have to remember this one too (pointing to card 3, then flip over card 4). Is this one the same as the one before?" Continue flipping over each card and asking the child yes or no.

“Now let’s do it on the computer. I’m going to do some first and you watch. I am going to press this button (point to the space bar) if the picture of the dots is the same as the picture that came two pictures before it.  Look, remember this. Ok, now we have to remember this one too. Now is this picture the same as the picture that came two before?” Repeat for each array and nod or shake your head accordingly.

“Ok, now it’s your turn to do it on the computer. I want you to press this button (point to the space bar) if the one you see is the same as the one that came two pictures before it. Just try your best. If you are not sure, it’s ok to guess.” Present the practice set on the computer. If the child does not get them all correct, repeat the practice again together: “That was close, let’s do it one last time together.”

“Good job!  Now you’re going to do this game three times by yourself.”  Complete the three trials. If a break is needed between trials, repeat the instructions: “Your job is to press this button (point to the space bar) if the picture of dots on the screen is the same as the picture that came two pictures before it.”

## Appendix H: Probe Sentences for the Referential Devices Pretest

| REFERENCE | PROBE WORDING |
|---|---|
| **My** | Ken said, "That is your bike and this is ___________." (my skateboard) |
| | Katie said, "This is your uniform and that is ___________." (my uniform) |
| **'s** | This is Holly. Whose bear is it? It is _______. (Holly's) |
| | This is Steve. Whose guitar is it? It is ______. (Steve's) |
| **Their** | This is his house and that is _______________. (their house) |
| | It is her present. It is not __________. (their present) |
| **His** | The boy has a football. The football is ______. (his) |
| | The boy is reading a book. The book is ______. (his) |
| **Hers** | The girl has a new rattle. The rattle is _____. (hers) |
| | The girl is putting on her new boots. The boots are ____. (hers) |
| **It** | Here is a bus. It is big. Here is a puppy. _________. (It is small) |
| | Here is a book. It is small. Here is a horse. ____________. (It is big) |
| **Them** | They have a new CD player to share. The CD player belongs all of ____. (them) |
| | The children are playing with a skipping rope. The skipping rope belongs to all of ___. (them) |
| **Him** | The boy has a lunch kit. The lunch kit belongs to _______. (him) |
| | The boy has a fishing rod. The fishing rod belongs to ______. (him) |
| **Her** | The girl has a book. The book belongs to ______. (her) |
| | She is opening a present. The present belongs to _____. (her) |
| **He** | She is eating a hamburger and _____________. (he is eating a hotdog) |
| | She is walking to school and ______________. (he is feeding the dog) |
| **She** | He is climbing and ___________. (she is sliding) |

| | |
|---|---|
| | He has a skipping rope and _________. (she has a ball) |
| **I** | Here they are skipping. Do you like skipping? ________. (I (don't) like it) |
| | Here she is swimming. Do you like swimming? _______. (I (don't) like it) |
| **Me** | Holly loves her bear. She says, "Look this teddy bear belongs to __." (me) |
| | Alex just got a present. It's a football. He says, "Really, this football is for ______?" (me) |
| **They** | The children are swinging. Who is swinging? ______. (They are) |
| | The people are running. Who is running? _______. (They are) |
| **You** | The dog looks hungry. The boy said, "Do ______?" (you want some food) |
| | Kirk said, "I want your hat!" Josh said, "No _______." (you can't have it). |
| | The dog wants to go outside. The boy asks, "Do______?" (you want to go for a walk) |
| **A** | Here is a grocery store. Here is ___________. (a school) |
| | The children are playing _____. (a game) |
| **The** | Here is a red bird. Here the red bird is eating a worm. Now, ________. (the red bird is flying) |
| | Here is a ladder. The man is climbing _________. (the ladder) |
| **Who** | Who is wearing red? The boy _______. (who is in a tree) |
| | Which baby is wearing purple? The baby _______. (who is in the stroller) |

## Appendix I: Word Lists for the Keep Track Task

| **Toys:** | **Transportation:** | **Clothing:** |
|---|---|---|
| Puzzle | Car | Shoes |
| Doll | Boat | Dress |
| Blocks | Bike | Hat |
| Kite | Bus | Belt |
| Balloon | Airplane | Shirt |
| Crayon | Truck | Socks |
| Book | Tractor | Pants |
| Ball | Train | Sweater |

| **Food:** | **Body Parts:** | **Animals:** |
|---|---|---|
| Apple | Arm | Cat |
| Bread | Ear | Dog |
| Cake | Eye | Frog |
| Carrot | Finger | Duck |
| Corn | Foot | Horse |
| Grapes | Hand | Turtle |
| Orange | Lips | Snake |
| Sandwich | Nose | Pig |