University of Alberta

Mapping Mazes: Developing a Taxonomy to Investigate Mazes in

Children's Stories

by

Amanda MacLurg

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Speech and Language Pathology

Department of Speech Pathology and Audiology

Spring 2014
Edmonton, Alberta

Dedication

This work is dedicated to Andrea, Shawna, Bernadette, and Nadine - women who persevered, sacrificed much, and poured their encouragement into me. This is as much yours as it is mine.

Abstract

Pausing, repeating, revising, and abandoning words and phrases are common characteristics of speech called mazes. Mazes affect the fluency of speech and are thought be related to language ability and language processes. Few relationships between mazes and language ability and language processes have been identified. This is due to methodological differences in how researchers have coded mazes and approached elicitation tasks. The main purpose of this study was to address these methodological challenges by developing a reliable and objective maze taxonomy. Once developed, the maze taxonomy was applied to the stories of Kindergarten and Grade Two students. Stories were elicited from oral and picture tasks to determine age or task effects on maze production. Overall, the Kindergarten and Grade Two students produced mazes at similar rates and used more mazes in stories elicited from oral tasks than picture tasks.

Acknowledgements

I would like to thank Dr. Phyllis Schneider for agreeing to supervise this project and for providing her direction, insight, and support. I would like to thank Dr. Joanne Volden for being a member of my thesis committee and sharing her expertise and perspective to make this project better. I would also like to thank my other committee member Dr. Paula Marentette for contributing her knowledge and advice during this endeavour, and for sparking my interest in research during my undergraduate degree. I would like to thank Dr. Carol Boliek for her technical expertise and Tawnya Schmidt for acting as the second rater for this study. I would like to acknowledge and thank the storytellers from Schneider and Dubé (2005) who contributed their stories to this project. My husband Mark sacrificed much to support me through this program. Thank you for your encouragement and confidence in my abilities. We are a good team. My family also dove into this adventure with me. Thank you for honouring the work that went into this program and this project. I would also like to acknowledge my two favourite storytellers Ana-Lee and Judah. Thank you for inspiring me to study stories. This research would not have been possible without the funding of Joseph Armand Bombardier Canada Graduate Scholarship from the Social Science and Humanities Research Council and the Walter H Johns Graduate Fellowship.

Table of Contents

List of Tables

List of Figures

Mapping Mazes: Developing a Taxonomy to Investigate Mazes in Children's Stories

Speech disruptions and repairs are a common occurrence in verbal communication. A person may pause, repeat, revise, or abandon a sentence seemingly unaware of the changes (Dollaghan & Campbell, 1992). Loban (1976) recognized the similarities of the speech disruptions/repairs to the behaviour of a human trapped in a maze (e.g., pausing, revising steps, and abandoning a path), and called the speech disruptions/repairs *mazes.* Mazes have been linked to language ability and language processes. For example, children with language impairments produce more mazes than their peers with typically developing language (Guo et al., 2008; Nettelbladt & Hansson, 1999; Thordardottir & Ellis Weismer, 2001).

It has been difficult to establish other clear relationships between mazes and language ability, or processes, due to the different coding procedures and the different tasks used between studies. Most maze taxonomies have been based on the assumed underlying processes associated with specific maze types. If applied, a more objective maze taxonomy would make it possible to identify other maze patterns across studies. The purpose of this study was to address these methodological coding challenges by constructing a maze taxonomy based on the description of the speech behaviours. Once developed, the maze taxonomy was applied to the stories of typically developing children in order to determine the reliability of the maze taxonomy. In addition, these data add to the limited maze research by investigating whether maze production differences exist between age groups and between different story elicitation tasks.

Mazes

There is no single label for the pauses, repetitions, revisions, and abandoned utterances that occur during speech. The label *maze* (Loban, 1979; Dollaghan & Campbell, 1992) is sometimes used to name the speech phenomenon because it is neutral with regards to the reason for or cause of the behaviour. Other labels, such as *communication breakdowns, speech disruptions*, *speech repairs*, *speech errors,* and *disfluencies* are used to name the phenomenon based on an inferred function, purpose, or process associated with it. For example, the label *speech error* links the phenomenon with planning, while the label *self-repairs* links the phenomenon to self-monitoring (Evans, 1985).

Other names for the phenomenon are based on the disruption of speech (e.g., *disfluencies*) (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001). The term *disfluency* accurately describes speech containing typical pauses, repetitions, revisions, and abandonments, but the term is often used interchangeably with stuttering. Nonstuttering-like disfluencies have similar characteristics to stuttering-like disfluencies; however, they are different. Nonstuttering-like disfluencies include polysyllabic-word repetitions, phrase repetitions, filled pauses (i.e., interjections), and revisions. Stuttering-like disfluencies include monosyllabic or part-word repetitions, prolongations (i.e., a lengthened sound), and blocks (i.e., prolonged air stoppages usually at the beginning of a word) (Zackheim & Conture, 2003). Starkweather (1987) suggested that the two disfluencies also differed in the underlying processes associated with them. Non-stuttering-like disfluencies are thought to represent linguistic fluency while stuttering-like disfluencies are thought to represent motoric fluency (i.e., coordination and execution of the speech processes). For the purpose of this paper, the speech phenomenon studied and labeled by other researchers as *communication breakdowns, speech disruptions*, *speech repairs*, and *disfluencies* are called mazes, as the term does not represent a bias towards a specific underlying cause or function and cannot be confused with stuttering.

The speech behaviour included under the maze umbrella described below is based on Dollaghan and Campbell's (1992) maze taxonomy. The main categories of Dollaghan and Campbell's taxonomy include: pauses, repetitions, revisions, and orphans. There are other maze classification systems; however, Dollaghan and Campbell's labels provide a description of the speech behaviour rather than an assumed function or purpose of the speech behaviour. For example, some maze types are labeled as *stalls* based on the stalling behaviour thought to be associated with them (Rispoli, Hadley, & Holt, 2008). Grouping maze types under a single heading and analyzing them together has made it difficult to compare results between studies. A classification system based on the characteristics of the mazes, rather than the function of the mazes, establishes an objective taxonomy and makes cross study comparisons possible. The classification of mazes used in this study and the proposed associated underlying factors are described below.

Pauses

Pauses are defined as silent or filled breaks between or within utterances (Dollaghan & Campbell, 1992). The pause subtypes include filled pauses, silent pauses, and pause strings. Filled pauses are verbal interjections within and between utterances that do not contain linguistic meaning. They are the *ums* and *ahs* (e.g., "(*Um*) and then Mr. Elephant was filled with anger"). Silent pauses are the non-verbal breaks within and between utterances that occur for at least a two second duration (e.g., "And one time she looked at the dinosaur *(:02.2)* bones"). Pause strings are multiple pause types in an utterance (e.g., "*(uh) (:03.4)* the guy looked at her").

There are different underlying factors associated with pauses. Silent pauses are thought to indicate language production challenges (Dollaghan & Campbell, 1992). Filled pauses are believed to be associated with word finding difficulties or used as a tool for retaining a turn in a conversation (Dollaghan & Campbell, 1992; MacLachlan & Chapman, 1988). Silent and filled pauses are also thought to be a stalling behaviour and have been labeled as one maze type called *stalls* by Rispoli et al. (2008). Rispoli et al. stated that stalls are indicative of glitches that arise during the planning stage of sentence production, such as grammatical encoding. Glitches, in Rispoli et al.'s estimation, arise before articulation occurs and result in the production of filled and silent pauses. Repetitions (described below) were also classified by Rispoli et al. as stalls.

Repetitions

Repetitions are the repetition of a sound, word, or phrase (Dollaghan & Campbell, 1992). The repetition subtypes in the Dollaghan and Campbell maze taxonomy include forward, partial, exact, and backward repetitions. Forward repetitions are repeated unfinished words or phrases, followed by the completion of the utterance (e.g., “and (*she*) she was happy cause she still had her pretty flower”). Partial repetitions are repeated unfinished words or phrases but the utterance is not completed (e.g., “(*And then*) and then”). Exact repetitions are the repetitions of sounds, words, or phrases that have already been completed (e.g., “(*She ran fast*) ran fast”). Backward repetitions are repetitions of the beginning of an utterance after a word or a phrase was inserted (e.g., “(*It wanted*), you know, it wanted a bone”).

The repetition subtype labels in Dollaghan and Campbell’s (1992) taxonomy are confusing. For example, the words or phrases in the forward and partial subtypes are repeated word for word, but they are not considered to be exact repetitions. The definition of backward repetitions is also unclear because it resembles a type of revision rather than a repetition. Word or phrases are repeated in backward repetitions but an alteration also occurs between the original word/phrase and the repeated portion. The alteration indicates that the maze was a revision. The repetition label discussed below and used in this study was based on Dollaghan and Campbell’s main repetition definition and not their repetition subtypes.

The repetition of words or phrases are thought to signify different underlying causes. Guo et al. (2008) suggested that phrase repetitions signify concept formulation difficulty or challenges in accessing particular syntactic structures. Word repetitions may arise from difficulties in retrieving word meanings. As mentioned above, repetitions are also thought to be another form of stalling behaviour (Rispoli et al., 2008).

Revisions

Revisions occur when part of an utterance is altered after it was initially delivered (Dollaghan & Campbell, 1992). They are further grouped into subtypes depending on the type of alteration that occurred. Revision subtypes include: alterations to fix errors (e.g., "Mr. Rhino was (*sitting*) was standing by the oranges"), to add information (e.g., "(*Mr. Rhino was very*) It was a big mess. Mr. Rhino was very mad."), or to delete information (e.g., "The hippo (*tripped the*) tried to get the bee"). Other revisions that did not fall into one of these three subtypes are labeled mystery revisions.

The label *revision* implies an underlying purpose for the maze type. The function of this maze type may be to: revise the intended message after giving the message more thought, revise the message after hearing it and comparing it with the intended message, or revise it after receiving feedback from the listener. Revisions are generally thought to be indicative of formulation difficulties and self- monitoring (MacLachlan & Chapman, 1988; Rispoli, 2003; Thordardottir & Ellis Weismer, 2001). Rispoli (2003) proposed that when a glitch arises in the grammatical encoding phase the speaker hears what was articulated, understands the message, and then compares the spoken message with the intended message. A revision may occur if the intended message and the spoken message do not align. Although the label revision may describe an underlying function, the label also reflects what the speaker is doing: the speaker is revising the message. For this reason the label *revisions* will be used in this paper, keeping in mind that the label is intended to represent the speech behaviour and that there may be other possible explanations or functions for this maze type.

Orphans

Orphans are the final maze type defined by Dollaghan and Campbell (1992). Orphans are the abandoned sounds, words, and phrases that have no relationship to the other units in an utterance. The subtypes of orphans are determined by the length of the orphan. Phoneme orphans are abandoned sounds (e.g., "Then (*sh*) it left the cage"), word orphans are abandoned words (e.g., "I wanted (*that*) my money") and string orphans occur if a sentence or phrase is abandoned (e.g., "(*and she*)").

Orphans are labeled by others as *abandoned utterances* or *false starts* (MacLachlan & Chapman, 1988). MacLachlan and Chapman (1988) suggested that orphans occur when an utterance is too complex for the speaker to complete, leading to abandonment of the utterance. Evans (1985) suggested that orphans indicate self-monitoring. The speaker hears what she is saying, compares it to the intended message, determines if it is correct or if it is still what she wants to communicate, and then may abandon the utterance to correct the message.

The mazes — pauses, repetitions, revisions, and orphans— discussed above have been classified according the taxonomy developed by Dollaghan and Campbell (1992). The reliability of the maze taxonomy was established by applying it to the language samples of individuals with and without brain injuries. Along with determining the reliability of the maze taxonomy, Dollaghan and Campbell sought to establish whether there were differences in the maze production between individuals with and without brain injuries. The participants ranged from 7 to 20 years of age. Half of the participants had brain injuries and half of the participants were their age-matched counterparts with no brain injuries. Conversational samples were elicited through a series of questions. The language samples were transcribed and mazes were coded into the various subtypes described above. Maze rate was calculated based on the total number of mazes per 100 non-mazed words.

Four of the twenty transcripts were coded separately by two trained examiners to establish reliability. Reliability was established with a main category agreement of 84% and subcategory agreement at 80%. No differences were found between the overall maze rates of the two groups. The individuals with and without brain injuries produced mazes at a similar rate. When maze types were analyzed, the results indicated that more silent pauses were used by the participants with brain injuries than the participants without brain injuries.

The reliability of Dollaghan and Campbell’s (1992) maze taxonomy was established and was applied to determine maze production differences (i.e., specifically silent pauses) between two populations. The Dollaghan and

Campbell taxonomy will be used below to describe and compare mazes between studies. As mentioned above, different studies have their own way of describing mazes depending on theoretical position and purpose of research. For clarity, the labels assigned to the mazes in each study will be described. Then the maze labels derived from Dollaghan and Campbell's taxonomy, which best align with the operational definitions of those labels, will be used to discuss the literature. The original labels will be placed in parentheses for reference.

Mazes and Language

Mazes have been studied to gain insight into language ability and language processes, particularly in children. Maze production has been observed in children as young as two years of age (Clark, 1978). Some have suggested that mazes reveal a child's linguistic competence and language level (Loban, 1976; Navarro-Ruiz & Rallo-Fabra, 2001). The link between mazes and language ability has been examined by comparing the maze production of children at different ages (Evans, 1985; Loban, 1976) and by comparing the maze production of peers with different language abilities (Guo, Tomblin, & Samelson, 2008; Loban, 1976; Thordardottir & Ellis Weismer, 2001). One consistent finding has been that children with lower language abilities use more mazes than their peers with typically developing language (Guo et al., 2008; Nettelbladt & Hansson, 1999; Thordardottir & Ellis Weismer, 2001). Other maze results differ between studies, particularly when comparing types of mazes. One explanation for the inconsistent findings is the methodological differences between studies, particularly in regard to maze definitions and elicitation tasks.

Language Ability

In clinical language assessments mazes are believed to indicate word finding and sentence formulation difficulties (Thordardottir & Ellis Weismer, 2001). It has been assumed by some that children with language impairments, or younger children whose language is just developing, will produce more mazes than children with a more developed language system (Loban, 1976; Nettelbladt & Hansson, 1999; Thordardottir & Ellis Weismer, 2001). In contrast, others believe that mazes indicate higher-level skills, such as self-monitoring, and therefore children with a more developed language system will produce more mazes (Evans, 1985). The only clear and consistent maze pattern found and reproduced across several studies was that children with language impairments produce more mazes than their peers with typically developing language (Guo et al., 2008; Nettelbladt & Hansson, 1999; Thordardottir & Ellis Weismer, 2001).

No other clear relationships were identified between maze production and language ability due to the difficulties in encountered by comparing the results between studies that used different maze coding procedures and elicitation tasks. For example, Loban (1976) coded and analyzed filled pauses, repetitions, revisions, and orphans all together. Silent pauses were not coded or analyzed. Thordardottir and Ellis Weismer (2002) coded and analyzed filled pauses separately from content mazes, which included repetitions, revisions and orphans. Silent pauses were also not coded. Guo, Tomblin, and Samelson (2008) coded and analyzed silent pauses separately from vocal hesitations, which included filled pauses, repetitions, revisions, and orphans. Silent pause durations were also compared. Though the majority of the studies coded for most maze types, the types were often one component of an overarching maze label (e.g., content mazes and vocal hesitations). The mazes under the overarching labels were analyzed as one entity and were assumed to display a common behaviour or underlying function. The different coding and analyses procedures have made identifying maze production patterns challenging.

Loban (1976) was one of the first to examine the relationship between maze production and language ability. The primary purpose of the longitudinal study, which followed participants from Kindergarten to Grade 12, was to identify obvious stages of language development. Mazes were one variable of interest and were thought to be useful in determining developmental differences in the flow of oral language. The participants were grouped by language proficiency into three groups: low language proficiency, random language proficiency (i.e., children across the language proficiency distribution who were randomly selected), and high language proficiency. Language samples were elicited annually through conversational interviews based on familiar topics (e.g., games, parties, and comics). The language samples were transcribed and mazes were coded along with many other variables. Unfortunately, no maze coding procedure was described, but assumptions about what was coded can be made based on Loban's maze definition:

> In many respects this behaviour in language resembles the physical behaviour of someone being trapped in a special maze, thrashing about in one direction or another, hesitating, making false starts, or needlessly

retracing steps, until finally they either abandon their goal or find the path (Loban, 1976, p. 10).

From the definition it can be assumed that Loban coded mazes as revisions (i.e., false starts), filled pauses (i.e., hesitation pauses), repetitions (i.e., retracing steps) and orphans (i.e., abandonments).

Loban (1976) used the mean number of words per maze to compare maze production of the participants. In retrospect, Loban stated that the method underrepresented the mazes produced by the low group. Based on the methodology for calculating maze rate it can be assumed that silent pauses were not coded because silent pauses do not have words attached to them. Loban's analysis counted maze words, not the presence of the maze.

Although statistical analysis was not conducted, the data suggest that the children in the high and random groups produced the fewest words per maze, while the children in the low group produced the most words per maze. Across participants, mazes were produced at similar rates in Kindergarten and Grade Twelve. The maze rates only changed during the middle school years, as age increased the number of words per maze between Grade Four and Grade Nine increased. The low group appeared to have the highest increase in words per maze during this period.

Loban (1976) sought to link oral language and maze production. One way that language level was determined was by the average number of words per communication unit (c-unit) produced. The average number of words per c-unit was calculated for each participant during each grade. The high group produced the most words per c-unit and the low group produced the fewest number of words per c-unit. All of the participants increased the number of words per c- unit as they progressed by grade. Loban claimed that the high group was more advanced in their oral language development than the low group, and asserted that maze production differences demonstrated this. Loban tied maze production to verbal planning. The low group had the most words per maze because it was more difficult for them to plan and formulate what they wanted to say. However, Loban acknowledged that maze production could not necessarily be predicted by language proficiency or vice versa. Occasionally there are children that produce the same proportion of mazes but have very different language abilities.

Loban (1976) was one of the first to look at the relationship between language ability and maze production. Other maze studies have built on Loban's work by modifying maze calculations to express maze rate by word or c-unit rather than calculating the number of words per maze. Comparing Loban's results with the results from other maze literature is challenging for a few reasons. First, maze production was calculated much differently from other maze studies. Second, no maze coding procedure was described. And finally, the language elicitation tasks are difficult to compare with other studies because they were unstructured conversational tasks. While conversations provide insight into the language abilities of an individual, conversations are spontaneous and flexible. Unlike storytelling or retelling, conversations do not have a specific format or a set of specific language elements that must be delivered. This makes it difficult to compare language samples between the participants and draw conclusions about the relationship between language ability and mazes.

Thordardottir and Ellis Weismer (2002) recognized the importance of elicitation tasks in studying maze production. Mazes are more likely to be produced in narratives than in conversations (MacLachlan & Chapman, 1988). Thordardottir and Ellis Weismer suggested that narratives require more complex language than conversations. Narrative elicitation tasks were chosen to obtain language samples for comparing maze production between children with and without language impairments. In the spontaneous narrative task the participants were prompted to talk about personal familiar experiences (i.e., their birthday party, a favourite movie, or a favourite book). The main objective of the study was to identify whether filled pauses and content mazes were produced differently between children with and without language impairments. Content mazes were defined as any maze that contained linguistic meaning.

The participants ranged in age from 5 to 10 years. Half of the participants had language impairments and half had typically developing language. The participants in each group were matched for age and a subset of participants were matched for mean length of utterance (MLU). It was predicted that the children with language impairments would produce more mazes than their MLU-matched counterparts because the children with language impairments were thought to have processing and working memory difficulties. No predictions were made about maze production differences between the children with language impairments and their age-matched counterparts.

Mazes were coded as filled pauses or content mazes. Filled pauses were described as the verbal non-linguistic material produced (e.g., *um* or *ah*). Content mazes included revisions, repetitions, and orphans (i.e., abandoned utterances). Content mazes were classified as one entity because all the mazes included linguistic material. Silent pauses were not coded. The total number of mazes in 50 utterances and the number of each maze type (filled pauses and content mazes) in 50 utterances were calculated.

The results indicated that the children with language impairments produced more content mazes than the MLU-matched group. The children with language impairments and the age-matched group produced content mazes at similar rates. The children with language impairments produced fewer filled pauses than the MLU-matched group and the age-matched group. Post hoc analysis revealed that the participants with typically developing language produced filled pauses and content mazes at similar rates. The children with language impairments produced more content mazes than filled pauses (Thordardottir & Ellis Weismer, 2002).

Thordardottier and Ellis Weismer (2002) stressed that different underlying factors are responsible for filled pause and content maze production and that the two maze types should not be analyzed together. Content maze production may be related to age. The older participants (the children with language impairments and the age-matched group) produced more content mazes than the younger MLU-matched group. Filled pause production does not appear to be related to age, as the filled pause production did not follow the same production pattern as the content maze production between the groups. Thordardottier and Ellis Weismer suggested that filled pause production may be tied to metalinguistic processes, but the hypothesis was not tested.

Lumping all the maze types that contain linguistic information under one heading (content mazes) does not reveal a complete picture of maze production. Repetitions, revisions, and orphans were coded and analyzed as one entity and it is possible that each maze type differs in the factors related to them. The results would likely differ if each maze type was coded and analyzed separately. Thordardottier and Ellis Weismer (2002) made an important contribution to the maze literature by highlighting the importance of analyzing maze types to understand the factors responsible for them, particularly when investigating maze production and language ability.

Nettelbladt and Hansson (1999) also investigated maze production and language ability by comparing the maze production of children with and without language impairments. The study tested the validity of the assumption that children with language impairments talk less and are less fluent then their peers. Ten Swedish pre-school children took part in the study. Half of the participants had language impairments and half of the participants had typically developing language and served as the MLU-matches. To obtain language samples, the examiner engaged the children in conversations using books, toy catalogues, and photo albums. The interaction was recorded and the first 10 minutes of the conversations were transcribed. Filled and silent pauses were coded under one heading (pauses), and repetitions and revisions were also coded according to the taxonomy specified by Dollaghan and Campbell (1992). The maze rate was determined by calculating the total number of mazes per 100 non-mazed words. Repetitions were categorized further as part-word or whole-word repetitions and described as either affecting a lexical word or a function word. Lexical words were defined as any noun, verb, adjective or adverb. Function words were defined any personal pronoun or pronominal adverb. The placement of mazes was determined as initial (i.e., mazes in the first part of the utterance) or non-initial (i.e., anywhere else in the utterance). Mean length of mazed utterances and the overall mean length of the utterances (MLU) were calculated and compared. Mazed words were excluded from the MLU calculations.

Nettelbladt and Hansson (1999) found that the children with language impairments produced significantly more mazes per 100 fluent words then the children with typically developing language. The children with language impairments produced significantly more pauses and repetitions than the children with typically developing language. Children with language impairments also produced more part-word repetitions than the children with typically developing language. There was no significant difference in the occurrence of repetitions in function or lexical words between the groups. Most of the mazes occurred in the initial part of the utterance although there was not a large difference in the repetition location.

Nettelbladt and Hansson (1999) suggested that the results supported the assumption that children with language impairments are less fluent than their peers. In particular children with language impairments produce more pauses and repetitions than their peers. Nettlebladt and Hansson demonstrated how specific maze types could be analyzed separately to link maze production to language ability. However, the results related to the specific maze types differed from the results of other studies. The study found that children with language impairments produced more pauses and repetitions than children with typically developing language. Guo et al. (2008) found that children with language impairments only produced more silent pauses than their peers. Thordardottier and Ellis Weismer (2002) found that children with language impairments produced fewer filled pauses than their age-matched counterparts with typically developing language. The maze type differences between studies could be due to the methodological differences across the studies such as the age of the participants, the coding procedures, and/or the nature of the task. Nettlebladt and Hansson also suggested that the results differed because the definition of a language impairment used may have differed from other studies.

Guo, Tomblin, and Samelson (2008) sought to address the methodological differences in previous maze studies. The study examined the relationship between mazes (i.e., speech disruptions) and language ability. Guo et al. proposed that glitches arise during language production which reflects the cognitive processes that underlie language processes. Mazes were thought to be the external representation of glitches and it was proposed that maze production would decrease as syntactic development occurred in children with typically developing language. Guo et al. proposed that children with language impairments would produce more mazes due to their lexical and syntactic deficits than their age-matched peers. It was also predicted that there would be no significant difference in maze production between the children with language impairments and their language-matched peers.

Like Thordardottir and Ellis Weismer (2002), Guo et al. (2008) explored whether language ability was a factor associated with maze production by matching the children with language impairments with children at similar language levels. The study predicted that children with language impairments would use more silent pauses and vocal hesitations than the age-matched group. Guo et al. reasoned that children with language impairments often require more time to formulate sentences so more glitches were thought to arise due to lexical and syntactic deficits. Guo et al. defined vocal hesitations as any maze that included verbal output. Filled pauses, repetitions, revisions, and orphans were coded and analyzed as one under the vocal hesitation label.

Grade Four students with language impairments were matched for language ability with Grade Two students with typically developing language. Two years later, the data from the Grade Four students with language impairments were matched with new data of the original Grade Two students for an age-matched group. Narrative language samples were collected from the participants in a story retell task where the participants were instructed to choose a story (i.e., a set of three pictures) and a scripted story about the pictures was read to the child. The children were asked to describe each picture and identify all the key elements in the story. If a participant missed a key element the examiner identified the missing key element. After this procedure the participants were instructed to tell the whole story using all three pictures.

The narratives were transcribed with SALT and coded into maze types: silent pauses (longer than 250 ms) and vocal hesitations. Silent pauses were categorized into different groups according to pause duration (e.g., 250- 500ms, 500-1000ms, etc.). Maze rate was calculated by dividing the total number of mazes by the total number of words in the sample. The total number of words, the number of communication units (C-units), and mean length of C-units in words and morphemes were also calculated for each narrative.

Guo et al. (2008) found that the Grade Four participants with language impairments produced fewer words overall than the age-matched group. The Grade Four participants with language impairments and the language-matched group produced a similar number of words. There were no significant differences found between the participants with language impairments and the age-matched group, or the language-matched group, for C-units, MLU in words, and MLU in morphemes.

The findings from Guo et al. (2008) suggest that there is a relationship between language ability and mazes. As was predicted, the participants with language impairments produced more mazes than the age-matched group and they produced mazes at a similar rate as their language-matched counterparts. The participants with language impairments produced more silent pauses than the language-matched group and the age-matched group, but not more vocal hesitations. When silent pauses were analyzed for duration, the participants with language impairments produced more silent pauses in the 500-1000 ms category than the age-matched group. There were no other differences found for silent pause duration between the groups. More silent pauses were produced than vocal hesitations across participants. Guo et al. analyzed the maze types under the vocal hesitation umbrella separately and did not find any significant differences.

The research design used in Guo et al. (2008) was strong. Children with language impairments were matched for language ability with children that had typically developing language. The children with typically developing language also served as the age-matched group two years later. Guo et al. is the only study that explicitly matched for language ability. The design is an example that other maze studies could use to further investigate the relationship between maze production and language ability. One challenge identified in the study was the maze coding and analysis of silent pauses. Guo et al. used a fairly liberal method for coding silent pauses by including any break greater than 250 ms as a silent pause. This differed from 2 second or greater duration required by Dollaghan and Campbell's (1992) maze taxonomy. It is difficult to compare the silent pause results between studies because of the duration differences.

In summary, Loban (1976), Thordardottir and Ellis Weismer (2001), Nettelbladt and Hansson (1999) and Guo et al. (2008) all examined the relationship between mazes and language ability and found that children with language impairments (or with lower language proficiency) produced more mazes than their peers with typically developing language. The results between the studies varied when maze types were compared. In Loban (1976) there appeared to be no difference in the types of mazes produced, as the participants got older, or between language proficiency groups. Thordardottir and Ellis Weismer (2001) found that the children with language impairments produced more content mazes than the MLU-matched group, and produced content mazes at similar rates as the age-matched group, suggesting that content maze production was related to age. The children with language impairments produced fewer filled pauses than the MLU-matched group and the age-matched group. In Nettelbladt and Hansson (1999) children with language impairments produced more pauses and repetitions than the children with typically developing language. Guo et al. (2008) found that children with language impairments produced more silent pauses that were 500-1000 ms in duration than their age and language-matched counterparts. It appears that a relationship does exist between mazes and language ability; however, it is unclear what maze types are related to language ability or to other factors due in part to the different maze coding procedures used between the studies.

Language Processes

Guo et al. (2008) touched on the interaction between mazes, language ability, and language processes as mentioned above. Guo et al. hypothesized that mazes are external representations the cognitive and language processes that occur during sentence production. Language ability and language processes are intertwined. While studying the relationship between mazes and language ability, the underlying language processes should also be considered. Mazes have been studied as potential indicators of self-monitoring (Evans, 1985; Rispoli, 2003), formulation difficulties (Guo et al., 2008; Rispoli, 2003) and encoding processes

(Kowal, O'Connell, & Sabin, 1975). The maze labels used in the studies (Evans, 1985; Rispoli, 2003; and Kowal et al., 1975) described below were based on the underlying process of which the mazes were thought to be indicative. While the coding procedures were appropriate for the specific research questions raised in each study, comparing the results and determining if other patterns exist is challenging.

Evans (1985) believed that mazes (i.e., speech repairs) were related to self-monitoring processes. Evans collected language samples from Kindergarten and Grade Two children during 'show and tell' activities in their classrooms throughout the school year. The language samples were transcribed and were coded for repetitions, revisions (i.e., corrections), revisions (i.e., postponements) and orphans (i.e., abandonments). Filled and silent pauses were not coded. Corrections and postponements were two maze types that were coded separately by Evans. According to the Dollaghan and Campbell (1992) taxonomy, the two types of mazes are essentially revisions. Corrections add information at the end of the utterance while postponements add information into the middle of an utterance. Maze rate was calculated by dividing the total number of mazes by the total number of words in each language sample.

The distribution of the maze types produced was similar across both groups. Across participants, the results indicated that more repetitions were produced than other maze types. The Grade Two participants produced more mazes than the Kindergarten participants. There were significantly more revisions (i.e., corrections) and orphans produced by the Grade Two students than the Kindergarten participants. Evans (1985) suggested the maze production differences between the Kindergarten and Grade Two participants was related to self-monitoring development. Evans hypothesized that the self-monitoring abilities of the older children were more developed so they used mazes to repair their speech errors. The younger children, with less developed self-monitoring abilities, did not recognize as many of their errors while they were talking and therefore required fewer mazes to repair their messages.

The maze coding and analysis used in Evans (1985) differed from other studies. Evans separated revisions into two types based on hypothesized function, corrections and postponements. Filled and silent pauses were not coded. It is likely that had corrections and postponements been analyzed as one (revisions), and filled and silent pauses were coded, the analysis would yield different results. If all maze types were coded, it would also be less challenging to use the results to determine if maze patterns exist across studies. Another significant challenge in Evans (1985) was that self-monitoring and mazes were not actually tested. Evans assumed that mazes were indicative of self-monitoring, but the task did not control for any variables related to self-monitoring. What was actually tested was the maze production rate between the age groups.

Like Evans (1985), Rispoli (2003) was interested in language processes and maze production. Rispoli proposed that grammatical and sentence changes that occur during language development would be represented through the type of mazes (Rispoli called these speech disruptions) produced in spontaneous speech. Rispoli hypothesized that some mazes, particularly those that did not add linguistic information, are used to buy time while the speaker works out sentence formulation problems. Other mazes that add linguistic information to the sentence are thought to be indicative of self-monitoring, as they try to convey the intended message to the listener. Rispoli classified all the mazes that were thought to ‘buy time’ as stalls. They included filled pauses, silent pauses, and repetitions. The mazes that were thought to change the utterance were classified as revisions.

Rispoli (2003) developed a theory of sentence and maze production based on previous models of sentence production. Three levels of sentence production were identified: grammatical encoding, phonological encoding, and articulation. The theory stated that sentences can be broken down into increments (e.g., if the sentence is “I always seem to be rushing” the first increment is “I always” and the second increment is “seem to be rushing”) (Rispoli, 2003, p. 820). For each increment, all three levels of sentence processing occur in a step-by-step fashion. Increments do not have to be completed before another increment can begin. Grammatical encoding (the first level) in the second increment can begin while articulation (the final level) is still occurring for the first increment.

Like Guo et al. (2008), Rispoli proposed that stalls arise when a glitch occurs at the grammatical encoding level. Glitches may result from challenges in constructing the appropriate element syntactically or difficulty with word finding. A glitch causes the flow of the sentence to be interrupted or slowed, including the remaining levels (phonological encoding and articulation), which can produce a filled pause, a silent pause, or a repetition (i.e., stall).

Rispoli hypothesized that revisions are more complex than the other mazes as they require self-monitoring. Self-monitoring requires two additional levels of processing: audition and comprehension. For example if a glitch arises in the grammatical encoding phase a person monitoring her speech will hear what she said (audition), understand what she said (comprehension), and then compare the spoken message with the intended message. If the messages are not congruent, revisions correct the message.

Rispoli (2003) proposed that revisions are very different from the other types of mazes (filled pauses, silent pause, and repetitions), and hypothesized that there would be a difference in the distribution of the different types of mazes within sentences. It was proposed that stalls would arise as a result of a glitch that occurred during an articulated increment (prospective in nature), rather than a revision that is articulated after an increment (retrospective in nature). Rispoli hypothesized that this meant that stalls would arise earlier in sentences than revisions.

Rispoli (2003) suggested that grammatical development was taxing for the child's language system, which would result in the production of stalls. It was predicted that as grammar developed and sentence production became less taxing for the child, stalls would decrease. It was also predicted that revisions would increase as a child's grammar developed; because the child had linguistic options available and self-monitoring abilities were maturing.

Rispoli (2003) tested these hypotheses. Language samples from natural conversations during play interactions were collected between participants (that ranged in age from 1;10 to 4;0) and their parents. The language samples were transcribed and were coded for revisions, repetitions, filled pauses, and silent pauses. Any pause that was between one to three seconds in duration was coded as a silent pause. Pause lengths were obtained through a stopwatch reading and the experimenter estimated the pause length of each pause over three separate stopwatch readings. Only pauses that were between one to three seconds on all three readings were coded as silent pauses. Unintelligible sentences, imitations, self-repetitions, or routine utterances (e.g., songs) were not transcribed. Utterance boundaries were determined by an unfilled pause longer than 3 seconds, more than two coordinate clauses, and/or appropriate end of sentence pitch fall.

Mean length of utterance (MLU) and the index of productive syntax score (IPSyn) were calculated for each participant. Each of the sentences included in the child's language sample was coded as being fluent, stalled, or revised. Stalled sentences included those with filled pauses, silent pauses; or repetitions of phonemes, syllables, whole word, and phrases. Any sentences that contained revisions were counted as revised sentences. If both a stall and a revision occurred in a sentence, the sentence was always counted as a revised sentence. Orphans were not coded. The location of the mazes in the sentence was also coded based on the number of words produced before the maze occurrence (e.g., the site was labeled 12 if there were 12 words before the maze).

Rispoli (2003) found that 15% of the sentences were stalled sentences and 6% of the sentences were revised sentences. It was also found that the 70% of the mazes were found in the first three words of the clauses. The stalls that occurred before the utterance were not included in the calculation. There were more stalls than revisions produced at the first word, although the relationship did not reach significance. Significantly more revisions were produced in the second word than stalls.

Stall production varied by child and was not related to MLU and IPSyn. There were significant positive linear correlations between revision rate and MLU, and revision rate and IPSyn. There was no relationship found between revision rate and age. There were no correlations found between stalls and MLU, IPSyn, or age. Rispoli (2003) concluded that the results support the hypothesis that revisions would increase as grammar develops, citing the relationship between revision rates and MLU, and revision rates and IPSyn. As children were able to produce longer sentences they produced more revisions. The study predicted that as grammar developed there would be fewer stalls produced, but the results did not support this prediction.

The coding and the analysis of mazes in Rispoli (2003) were based on a theory of sentence development and maze production. Filled pauses, silent pauses, and repetitions were all coded and analyzed as one type of maze (stalls) according to Rispoli's theory. The coding procedure was appropriate for Rispoli's study, but it is difficult to compare Rispoli's results with others that investigated mazes and language processes.

Kowal, O'Connell, and Sabin (1975) also examined the relationship between mazes and language processes, more specifically formulation and encoding processes. Kowal et al. proposed that mazes (i.e., hesitations) were indicative of the encoding process related to message formulation. The primary purpose of the study was to collect normative data on the developmental trends of temporal patterning and mazes. The temporal patterning aspect of the study will not be discussed as it does not pertain to this study.

Language samples were obtained from participants with typically developing language that ranged in grade from Kindergarten to Grade 12. Each age group was comprised of 12 boys and 12 girls. The participants were presented a story generation tasks to elicit language samples. The task consisted of nine Snoopy cartoon pictures. Participants were asked to sequence and tell the story however they preferred. Half of the participants in each group told their stories individually to the two experimenters and half told their stories to an experimenter and a peer. In the latter condition the partner pair sequenced the story together and took turns acting as the peer listener and the storyteller. In the other condition the participants sequenced the story independently. The language samples were recorded and transcribed. Mazes were coded as silent pauses (270 ms or greater), filled pauses, repetitions, revisions (i.e., false starts), and parenthetical remarks. Parenthetical remarks were defined as verbal fillers that did not contain linguistic information such as ‘well’, ‘you know’, and ‘sort of’. The total number of each maze type by grade was determined.

The two environmental conditions (i.e., the presence of only the experimenters vs. a experimenter and the presence and aid of a peer) did not influence maze production. Kowal et al. (1975) reported the total number of each maze type produced by each group, although maze rate was not reported. The raw maze production totals indicated that 69% of the mazes produced were silent pauses, 9% were revisions, 8.4% were parenthetical remarks, 7.9% were filled pauses, and 5.7% were repetitions. Maze type production was not compared between groups and the raw data of each maze type by grade did not reflect the trends that Kowal et al. reported.

Kowal et al. (1975) reported that filled pauses and revisions remained relatively stable across grade levels. Repetitions appeared to decrease as the grade level increased. In contrast, parenthetical remarks appeared to increase as grade level increased. The silent pause trend was not described. Kowal et al. (1975) demonstrated that there may be different maze production trends as age or grade level increases. However, the purpose of the study was to relate mazes to encoding processes and the association was not found. Kowal et al. cited the low maze frequency as the cause for the lack of association. Kowal et al. (1975) suggested that a conversational format might yield more mazes due to a higher level of speaking demands and self-monitoring. However, this suggestion cannot be supported, given that later studies found more mazes in narratives than conversations (MacLachlan & Chapman, 1989; Navarro-Ruiz & Rallo-Fabra, 2001). Kowal et al. coded for all main maze types; however the analysis of the mazes was unconventional. It is possible that there were few mazes produced and using statistical analysis was not appropriate. The results and the trends that were reported provide a good basis for other maze research, but comparing the results with other studies and identifying common patterns is problematic.

In summary, in the studies that investigated maze production, language ability, and language processes, the one consistent finding was that children with language impairments produce more mazes than children with typically developing language (Guo et al., 2008; Nettelbladt & Hansson, 1999; Loban, 1976; Thordardottir & Ellis Weismer, 2001). Other results were inconsistent across the maze studies. Methodological differences are the likely cause of inconsistent results. The main methodological challenge identified was the different maze coding and analysis procedures used across the studies. Comparing and identifying maze patterns is challenging when some maze types were coded and analyzed while others were not. Different maze types have been analyzed under one overarching label (e.g., vocal hesitations) and compared with other maze types. Researchers have recognized this methodological problem and have called for a unified approach to coding mazes (Dollaghan & Campbell, 1992; Guo et al., 2008; Rispoli et al., 2008).

Maze Production in Different Elicitation Tasks

Another methodological challenge in interpreting the maze results across studies has been the different elicitation tasks that were used. Elicitation tasks influence maze production as evidenced by MacLachlan and Chapman (1988), Navarro-Ruiz and Rallo-Fabra (2001) and Schneider (1996). Studies have explored maze production between different task types and found that more mazes were produced in narratives than in conversations (MacLachlan & Chapman, 1988; Navarro-Ruiz & Rallo-Fabra, 2001). It was also found that children with language impairments produced more mazes in stories elicited from oral tasks than picture tasks (Schneider, 1996). Interpreting and comparing the results across the studies would be easier if the same terms were used to describe the elicitation tasks and the tasks were controlled to investigate specific hypothesis or predictions.

MacLachlan and Chapman (1988) used narrative and conversational elicitation tasks to determine whether sentence formulation was more challenging for children with language impairments than their peers with typically developing language. MacLachlan and Chapman hypothesized that mazes were indications of formulation difficulties such as word-finding difficulties. Specifically, they proposed that if wording-finding difficulties existed, more filled pauses and repetitions would be produced. They also predicted that children with language impairments would produce more revisions than the children with typically developing language. Their reasoning was that children with language impairments were more likely to make mistakes and would require more revisions to repair their speech than their peers. MacLachlan and Chapman also predicted that more mazes would be produced in narratives than in conversations. Narratives are believed to have higher syntactic demand, less listener support, and a specific organizational structure, supporting the belief that narratives were more challenging than conversations.

The children with language impairments (ages 9-12) were age-matched with children with typically developing language and were also matched for mean length of communication unit (MLCU) with younger children (ages 3-5) with typically developing language. There were seven participants in each group.

Conversation samples were obtained through conversational interviews based on familiar topics (e.g., family, hobbies, and events) and narrative samples were obtained by asking each participant to talk about a familiar TV show or movie. The language samples were transcribed. The number of communication units (c-units) and the mean length of sentences were determined. Mazes were coded as filled pauses and repetitions (i.e., stalls), revisions (i.e., repairs), orphans (i.e., abandoned utterances) and other. The maze rate was calculated by dividing the total number of mazes by the total number of c-units.

The results indicated that more mazes were produced in narratives than in conversations across participants. The maze production difference between the two tasks was significantly higher for the children with language impairments than the age-matched and MLCU-matched groups. MacLachlan and Chapman suggested a contributing factor to the higher maze rate produced in the narratives was MLCU, as overall MLCU was longer in narratives than conversations. More mazes were produced in utterances with longer c-units (10-12 words) than in utterances with shorter c-units (1-3 words). It was hypothesized that the maze production differences between narrative and conversational tasks were higher for the children with language impairments because their sentences were longer in narratives than in conversations and they had limited language skills, which lead to more formulation problems.

MacLachlan and Chapman (1988) found that filled pauses and repetitions (analyzed together as stalls), revisions, and orphans were produced similarly across groups. The hypotheses that connected specific maze types to specific processes were not supported. Stalls were later separated into filled pauses and repetitions and analyzed separately. The results indicated that the children with language impairments and the age-matched group produced more filled pauses than the MLCU-matched group. MacLachlan and Chapman stated that the filled pause results indicated that filled pause production increased with age.

MacLachlan and Chapman (1988) were among of the first to compare maze production between two different tasks by recognizing that the task demands between conversational and narrative tasks might differ. The narrative tasks that were used were unstructured and were conversation-like. They differed from the conversation tasks only because they included a story component. The participants were asked to describe an episode of a TV show or a movie, which typically contain a story and story elements. This comparison suggests that the task effects are strong and that there are specific elements in narratives and conversations that influence maze production.

Another study that compared maze production between two elicitation tasks was Navarro-Ruiz and Rallo-Fabra (2001), who sought to establish whether there were differences in the type of mazes produced by children with language impairments and children with typically developing language. Mazes were classified into three main types: mazes related to fluency, mazes related to turn taking, and mazes related to self-repair or unfinished clauses. The mazes that were related to fluency included: silent pauses, repetitions, and hesitations. No definition was provided for hesitations. The mazes that were related to turn taking included: filled pauses and discourse occupation markers. No definition was provided for discourse occupation markers. The mazes that were related to self-repair or unfinished clauses included: revisions and orphans (i.e., abandoned utterances).

The participants in the study were 6 - 8 years of age. Four children with language impairments were age-matched with four children with typically developing language. Language samples were obtained through a series of three tasks: a conversational task, a storybook retell task with the storybook as a reference, and an animated-cartoon retell task. Language samples were transcribed with SALT and mazes were coded into the three maze types described above.

Although no statistical analysis was reported, Navarro-Ruiz and Rallo-Fabra (2001) stated that children produced more mazes in narratives than in conversations. The participants with language impairments produced more filled pauses in conversations than narratives. In contrast, the children with typically developing language produced more filled pauses in narratives than in conversations. Across participants there were more filled pauses and repetitions produced in narratives than in conversations. The authors suggested that this occurrence was due to planning and memory processes. The participants with language impairments produced more closed-class word repetitions than the participants with typical language and the participants with typical language produced more filled pauses, revisions, and orphans than the participants with language impairments. Navarro-Ruiz and Rallo-Fabra (2001) suggested that the participants with typically developing language used more revisions and orphans because their error detection skills were better than the participants with language impairments.

Apart from the major challenges related to the maze coding methods, significant difficulties arose in interpreting the maze production results between the three tasks used in the study (Navarro-Ruiz & Rallo-Fabra, 2001). Maze results from the comparison between the two narrative tasks were not provided, nor were the maze statistics or data from the comparison between the conversation and narrative tasks. It is possible that a statistical analysis was not possible due to the small number of participants in the study, particularly because there were so many variables in the study (i.e., maze subtypes, age, and language level of participants). A larger sample size, a clearer definition of the maze types used, and a full report of the analysis and the results would have been beneficial for comparing these results with other studies and making claims about maze production as whole.

Wagner, Nettlebladt, Sahlen, and Nilholm (2000) also used narrative tasks and conversational to investigate whether there were task differences for intelligibility, fluency, and MLU for pre-school children with language impairments. The study also sought to determine whether there were syntactic and grammatical differences between the tasks. Language samples were gathered from 28 Swedish preschool children with language impairments (with a mean age of 5.4) through two conversational tasks and two narrative tasks. In conversational task one, the participants talked about topics of interest and in conversational task two, the participants answered interview style questions. In narrative task one, the participants were told fables and were asked to retell them and in narrative task two, the participants were shown pictures and were asked to tell a story about the pictures. The samples were transcribed and coded according to the SALT protocol. Mazes were coded as filled pauses, repetitions, and revisions. Fluency was coded as the proportion of complete and intelligible utterances that contained mazes. The samples were coded for intelligibility, MLU, and grammatical variables.

Wagner et al. (2000) found a larger number of complete and intelligible utterances in conversations than in narratives, and a higher number of mazes and higher MLU in narratives than in conversations. Mazes were counted and analyzed together. No analysis was reported on the individual mazes types. The narrative samples obtained were too brief to be compared statistically so all of the narratives were collapsed and analyzed together. The two types of conversational samples were also collapsed and analyzed together. It is unfortunate that the language samples from the narrative conditions were too brief to compare statistically as it is possible that the two narrative tasks and the perhaps the two conversation tasks would provide more information on how maze production differs between types of tasks.

Defining a Story

Controlling the type of language elicitation task used while studying mazes is important because maze production varies depending on the task. Different narrative tasks influence mazes differently (Schneider, 1996), supporting the notion that not all narrative tasks are alike and hence cannot be compared as one type of condition. Controlling narrative elicitation tasks allows for stronger comparisons. The methods of the narrative tasks used in maze studies differ. For example some narrative tasks are story construction tasks based on pictures (Kowal et al., 1975) while others are more conversational (MacLachlan & Chapman, 1988; Thordardottir & Ellis Weismer, 2002). One way to control narrative tasks and have strong comparison conditions is to control for story characteristics through story grammar units.

Stein and Glenn (1979) identified story grammar as the main element that set narratives apart from other language tasks. Stein and Glenn suggested that a good story includes information about the setting, an initiating event (i.e., the conflict or dilemma), an internal response (i.e., a character's reaction to the event), an attempt (i.e., a plan to solve the problem), a consequence (i.e., what happened as result of the attempt) and a reaction (i.e., how the character felt). Like the narratives in Schneider (1996) the narrative tasks in this study were designed around Stein and Glenn's story grammar model. The narratives that were elicited in this study have been called *stories* in order to distinguish them from other narrative tasks that have been used in the past that may not be congruent with Stein and Glenn's definition of narratives.

Schneider (1996) used a controlled story elicitation task based on Stein and Glenn (1979) to determine how children with language impairments tell stories when presented with different narrative elicitation tasks. Schneider (1996) found that children produced mazes at different rates depending on the type of story elicitation task. Sixteen children with language impairments aged 5 – 9 took part in four story retell conditions. In the pictures-only condition each child was presented with a set of pictures and was asked to tell the story about the pictures. In the oral-only condition the examiner told a story to the child, a listener was brought into the room and the child retold the story to the listener. In the oral-followed-by-pictures condition the researcher told the child the story and then pictures were laid out in front of the child. A listener was brought in and the child retold the story to the listener. In the oral-with-pictures condition the child was presented with the pictures, the researcher told the child the story, a listener was brought into the room, and then child retold the story to the listener. Four different stories were used for the elicitation tasks, so the children never told each story more than once. The story condition combinations were counterbalanced. Story grammar units were controlled in the presentation of all the stories. In the oral stories, each story grammar unit was included once. In the picture stories, each story grammar unit was included once in the pictures. It was up to the participants to identify and construct the stories with the story grammar units illustrated in the pictures.

The stories were transcribed and coded according to SALT. The stories were coded for mazes, which included repetitions, revisions, orphans (i.e., false starts) and filled pauses, but the type of maze was not indicated or analyzed separately. Mean length of T-units (i.e., average length of sentences), number of words, number of utterances, and the number of mazes were determined. The presence of story grammar units, the number of story grammar units, the number of different/relevant story grammar units, and the number of different/irrelevant story grammar units, cohesion rate, and cohesion errors were also coded by the researchers.

The results from Schneider (1996) that were pertinent for this study were the number of mazes that were produced in each condition. Schneider was interested in comparing maze production across conditions to investigate the hypothesis that mazes were linked to message-formulation load. The total number of mazes was compared across conditions and it was found that more mazes were produced in the oral condition and the oral-with-pictures condition than in the picture condition. The stories with an oral component were better, more complex stories (i.e., contained more story grammar units) than the stories elicited with just pictures.

Schneider suggested that these differences occurred because the children had more formulation problems in the conditions that contained an oral component than in the picture conditions. The formulation problems were believed to be related to memory load and the complexity of the story. In the stories with an oral component, someone else created the story and the storyteller was required to store that information, even with a picture in front of them, they retold the story the same way it was verbally presented. In the pictures-only task the storytellers made up the story and did not have to recall specific details about the story. They also could use the pictures to support the story that they created. The memory load is higher in the stories with an oral component.

In a similar study, Schneider and Dubé (2005) sought to determine how Kindergarten and Grade Two children with typically developing language told stories when presented with different story elicitation tasks. Schneider and Dubé followed the elicitation protocol from Schneider (1996), but mazes were not coded and analyzed, as they were not pertinent for answering the main research questions of the study. Schneider and Dubé found that across participants, better stories (i.e., more story grammar units) were produced when elicited from a story tasks with an oral component than a pictures-only story task. The stories used in this study (described below) were previously collected stories from Schneider and Dubé (2005).

In summary, one challenge in interpreting the maze results across studies has been the different elicitation tasks that were used. Elicitation tasks influence maze production. Children produce more mazes in narratives than in conversations (MacLachlan & Chapman, 1988). Children with language impairments produced more mazes in stories elicited from oral tasks than picture tasks (Schneider, 1996). One way to address the maze interpretation challenges across studies is to clearly define the task types. Controlling the tasks for specific elements also allows for stronger comparisons between task types. The story tasks used in this study and discussed below have adopted these methods.

This Study

Comparing and identifying maze production patterns across studies is challenging due to the methodological differences between the studies. This study has addressed the methodological differences by establishing a more objective maze taxonomy based on maze behaviours rather than the hypothetical functions of the mazes. The initial story presentation in each elicitation task used in this study controlled for story characteristics, allowing for a stronger maze production comparison between the tasks. Two participant groups of differing ages (Kindergarten and Grade Two) were used in this study to add to maze data regarding age and maze production differences.

Maze Taxonomy

The maze taxonomy established in this study was based on the taxonomy developed by Dollaghan and Campbell (1992). The Dollaghan and Campbell taxonomy was chosen as the framework for the proposed maze taxonomy because the definitions and labels describe the behaviour of the mazes rather than the function of the mazes. This more objective approach can be adopted by researchers with differing viewpoints. Dollaghan and Campbell's (1992) main maze categories, including: repetitions, revisions, and orphans were adhered to. The taxonomy in this study diverged from the Dollaghan and Campbell taxonomy by coding filled pauses and silent pauses as separate core maze types, restricting the mazes coded as repetitions, altering the definition of orphans, and omitting the subtype classifications. The maze coding taxonomy used in this study, along with examples, can be found in Appendix A. The main coding definitions of each maze type and the deviations from the Dollaghan and Campbell taxonomy are described below.

Pauses

In this study filled pause and silent pauses were coded as distinct mazes. Filled pauses were defined as non-lexical, one-syllable fillers (e.g., ‘”um’” and “ah”) and silent pauses were defined as any break in the utterance, 2 seconds or greater. The 2 second silent pause duration was used in this study to ensure that no natural pause breaks were mistaken as silent pauses as the majority of silent pauses produced are less than 2 seconds, many of which are thought to be natural pause breaks (Dollaghan & Campbell, 1992). The filled and silent pause coding and analysis differs slightly from Dollaghan and Campbell (1992). Dollaghan and Campbell collapsed filled and silent pauses under one main maze label (pauses). Nettelbladt and Hansson (1999) was the only other maze study that followed this protocol because they coded mazes according to the Dollaghan and Campbell taxonomy. Other studies treated filled and silent pauses differently by coding both pause types separately, or they coded only one pause type and did not code the other pause type (Guo et al., 2008; Loban, 1976; MacLachlan & Chapman, 1988; Thordardottir & Ellis Weismer, 2002; Wagner et al., 2000). Filled and silent pauses are different speech behaviours as filled pauses produce verbal output (e.g., “ah” and “um”) while silent pauses do not. For the reasons stated, in this study filled and silent pauses were coded as separate maze types.

Revisions

In this study any addition, substitution, or deletion of a portion or whole utterance was labeled as a revision following Dollaghan and Campbell's (1992) revision definition.

Repetitions

In this study, repetitions were defined as the exact repetition of sounds, words, or phrases based on Dollaghan and Campbell's (1992) definition. Although subtypes were not coded in this study, a clear distinction will be made between the types of mazes Dollaghan and Campbell considered to be repetitions and the types of mazes labeled repetitions in the maze taxonomy of this study. In this study only utterances that were repeated exactly, with no alteration to the beginning or middle of the utterance, were coded as repetitions. Dollaghan and Campbell included backward repetitions (i.e., a word or a phrase added to an utterance and the beginning of the utterance is repeated) as a repetition subtype. The definition of backward repetitions suggests that backward repetitions could also be considered revisions. For purposes of clarity, only exact repetitions with no alteration to the beginning or middle of the utterance were coded as repetitions in this study.

Orphans

The orphan definition used in the maze taxonomy of this study differs from Dollaghan and Campbell's (1992) definition that described orphans as abandoned sounds, words, or phrases were not related to the other units within an utterance. It is difficult to reliably distinguish orphans from revisions based on

Dollaghan and Campbell's definition, as no information is provided on how to determine whether or not a revised segment is related to the rest of the sentence. The taxonomy used in this study addressed subjectivity involved in coding orphans by describing orphans and setting the following coding boundary. Orphans were defined as abandoned utterances followed by a silent pause or a new story/conversation. If the coder had any doubt that the original utterance was or was not related to the linguistic information that followed it, the maze should be coded as a revision.

Subtypes

No maze subtypes were included in the maze taxonomy of this study. Comparing the maze subtypes would be challenging statistically. Dollaghan and Campbell (1992) completed a statistical analysis on the maze subtypes and found that people with brain injuries produced significantly more silent pauses. No other significant subtype differences were found. In this study, silent pauses are not considered a subtype and there is no other evidence to suggest that the other maze subtypes identified should be coded.

Story Elicitation Tasks

In order to compare language samples and elements of language samples (i.e., mazes) it is important to consider the elicitation task methods. One difficulty in comparing mazes between studies has been that language samples were elicited in a variety of ways. Maze research would benefit from moving toward using more controlled methods for eliciting language samples. This would provide more clarity in comparing mazes between studies because the tasks would be structured to account for confounding variables.

The stories used in this study were obtained from a previous study (Schneider & Dubé, 1997, 2005) that controlled the initial story presentation of each task. Three story elicitation tasks were used: oral, picture, and oral-picture. Only the stories elicited from the oral and picture tasks were used in this study. The storytelling tasks used in this study differed from the narrative tasks used in other studies (Kowal et al., 1975; MacLachlan & Chapman, 1988; Thordardottir & Ellis Weismer, 2001). In this study, during the initial presentation of the stories in the oral and picture tasks, each story grammar unit (i.e., initiating event, internal response, attempt, consequence, and reaction) was included once. The presentation of the stories in each elicitation task was controlled. Other studies did not control the story characteristics in the initial presentation or content around which the participants based their stories. For example, in the narrative task in Thordardottir and Ellis Weismer (2001) participants described personal familiar experiences (e.g., birthday party, favourite movie, or favourite book).

Controlling for the story characteristics in the initial presentation did not mean that the actual elicitation task was completely controlled. The actual stories or narratives that were produced in this study, and the story characteristics included, were ultimately determined by the participants. However, controlling the story characteristics in the presentation of each elicitation task does provide a stronger method for comparing stories and mazes within stories than had the story characteristics in the presentation tasks not been controlled at all.

A methodological concern arose while considering the task construction and maze differences. In the picture task there is possibility that more pauses occur between utterances as children move from one picture to another due to natural pause breaks. If this was the case, silent pauses in the picture task could be inflated. No other study has identified this potential methodological challenge. For that reason, in addition to the silent pause coding procedure outlined in Appendix A, silent pauses were also coded by location: between utterances [BSP] or within utterances [WSP]. If natural pause breaks occurred between pictures and inflated silent pause rates, silent pauses between utterances would be higher in the picture task than in the oral task.

Age

Some studies (Evans, 1985; Kowal et al., 1975; Loban, 1976; Rispoli, 2003) have tried to establish a relationship between mazes, language ability, and language processes by comparing the maze production of children of different ages. While some studies have not been able to link age to maze production (Kowal et al., 1975; Loban, 1976), one study (Evans, 1985) found evidence that age is related to maze production, as older children (Grade Two students) produced more mazes than younger children (Kindergarten students). The groups used in the current study were the same ages as the groups used in Evans (1985). In this study, the maze production of Kindergarten and Grade Two students was compared to add to the research in this area, seeking to determine whether mazes were produced differently by children of different ages.

Objectives

The objectives of this study were as follows:

1. The priority of this study was to provide a reliable maze taxonomy.
2. Based on the maze literature, the effect of age and task on overall maze production and specific maze types were examined.

Methodology

Participants

The participants in this study were obtained from a previous study that involved determining whether presentation effects influenced children's use of referring expressions and the content in their retelling of stories (Schneider & Dubé, 1997, 2005). Forty-four typically developing, English-speaking children from Edmonton in two grades took part in the study. Of the 22 participants in the Kindergarten group, 9 were male and 13 were female, with a mean age of 5.58 years (*SD*= 0.36, range 5.02 – 6.10). Of the 22 participants in the Grade 2 group, 10 were male and 12 were female, with a mean age of 7.81 years (*SD*= 0.36, range 7.08 – 8.61).

Materials

Three stories were used as the stimuli in two story elicitation tasks. There was an oral and picture version of each of the three stories. In all three stories there was a main character that was a female hippopotamus and a secondary character that differed for each story. The picture version of each story consisted of 5 images taken from Mercer Meyer’s book *Oops* (1977). The oral version of each story was developed based on the pictures. Story grammar units (i.e., setting, initiating event, internal plan, internal response, attempt, consequence, and reaction) (Stein & Glenn, 1979) were included once in the presentation of each story in both the oral and picture conditions. The units were included verbally in the oral condition; in the picture condition, the pictures depicted the units visually, or in the case of internal elements, depicted facial expressions that suggested the characters' thoughts and feelings.

Procedure

Each participant was asked to retell the stories over two story elicitation conditions: the oral condition and the picture condition. The story presentation order and the elicitation task order were counterbalanced. (1) In the oral condition the examiner told the story to the participant without showing the participant pictures of the story. (2) In the picture condition the examiner laid the pictures on the table in front of the participant and instructed the participant to look at the pictures. Each participant was allowed to take as long as he or she wanted to look at the pictures. The participant was then asked to tell the examiner the story.

After the presentation of each story, a listener was brought into the room. The participant had been told that the listener had neither seen nor heard the story and participant was to ‘tell the story as clearly as possible’ to the listener. In the conditions involving pictures, the pictures remained on the table and the child was allowed to look at them while telling the story. The listener was instructed to use neutral responses during the storytelling such as “uh huh”, “oh”, and “okay”. If there was no obvious ending to the participant’s story the listener was permitted to ask the participant if that was the end.

Transcription and Coding

All of the stories were recorded on videotape and audiotape. The stories were transcribed from the audio recording using SALT- Systematic Analysis of Language Transcripts (Miller, 2002) and checked while viewing the videotape.

The stories had been previously coded for mazes, which included repetitions, revisions, orphans (i.e., false starts) and filled pauses, but the maze types were not indicated. Only the mazes produced while the participant was telling the stories were coded. Mazes that occurred during participant-examiner conversations were not coded. Mazes were recoded in the current study with the maze taxonomy described in Appendix A. The primary maze categories included: filled pauses, silent pauses, repetitions, revisions, and orphans. The location of each silent pause was also coded as between [BSP] or within [WSP] the utterance (e.g., "(Um) [FP] (:02) [BSP] the (:02) [WSP] oranges came crashing down"). The duration of each silent pause was determined using the Audacity Software.

Multiple and successive filled pauses and repetitions were coded and counted in one parenthesis (e.g., "One day (ah ah) [FP] (sh she) [RP] she was putting on the train tracks"). Multiple revisions and successive revisions were coded and counted in separate parenthesis (e.g., "(She w) [RV] (then the guy said) [RV] and then she walked away"). The overall maze rate and the maze rate for each maze type were calculated by dividing the total number of mazes in each story by the total number of words in each story. Words produced during participant-examiner discussions were not included in the word totals. Words in the mazes were also not included in the word totals following the methods of Dollaghan and Campbell (1992).

The recordings for both the oral and picture stories of one Grade Two participant were damaged. The stories were transcribed before the recordings were damaged and mazes were included in the transcription. The examiner from this study coded all the mazes in the transcripts according to the procedures of this study, except silent pauses. Silent pause duration could not be accurately determined and could not be completed. In order to include the participant's other data in the overall maze rate analysis, the Grade Two mean silent pause rate was used for this participant's score. In the analysis of specific maze types, the silent pause rates for this participant were left out.

Overall maze rate was calculated by dividing the total number of mazes in each story by the total number of non-mazed words. The rate of each maze type was determined by dividing the total number of mazes of each specific maze type by the total number of non-mazed words.

Reliability

After the stories were initially coded, a test of inter-rater reliability was performed. A speech and language pathology graduate student acted as the second rater in this study. The second rater was trained to code the mazes using the maze taxonomy described above and found in Appendix A. She was instructed to code only the mazes within the story. In the event that child communicated with the experimenter, mazes were coded before but not during the conversation. If the story continued after the conversation, coding continued. The raters coded five practice transcripts together and five practice transcripts separately. The practice transcripts were obtained from the oral-picture condition that was omitted from this study. The coding for the practice transcripts was compared and any differences were discussed.

Eighteen stories (i.e., 20% of the transcripts) were randomly chosen and assigned a number to blind the second rater to the group and the condition. Of the stories chosen, four were from the Kindergarten oral condition, five were from the Kindergarten picture condition, four were from the Grade Two oral condition, and five were from the Grade Two picture condition.

Data Analysis

An analysis of inter-rater reliability was performed in order to establish whether the maze taxonomy developed by this study was reliable. A paired samples t-test was performed to determine whether there was a difference in silent pause production between utterances between the oral and picture tasks. To determine whether maze production differences existed between age groups and between elicitation tasks, a 2x2x4 mixed ANOVA was completed. The variables included age (2 levels), elicitation task (2 levels), and maze type (4 levels).

Results

The goals for this study were to establish a reliable method to code mazes, and to apply the constructed maze taxonomy to determine whether children in Kindergarten and Grade Two produced mazes differently when telling stories elicited from two different storytelling tasks. An analysis was performed to determine whether maze production differences existed between the age groups, between elicitation tasks, and between the age groups in each elicitation task. Specific maze types were also analyzed to determine whether they were produced differently across groups and tasks, between age groups, between elicitation task, and between age groups in each task. An exploratory analysis of silent pause location was also performed. The reliability of maze taxonomy, the silent pauses location analysis, and maze production between the age groups in both elicitation tasks are described below.

Reliability

The inter-rater reliability for coding mazes with the maze taxonomy developed in this study was calculated using Cohen's Kappa. The raters were 93.8% in agreement, or "almost perfect" according to Cohen's Kappa Index (Olmos, 2007). A detailed description of inter-rater agreement can be found in Appendix B.

Silent Pause Location

A methodological concern was raised regarding the occurrence of silent pause breaks that may occur between the utterances in the picture task. Paired sample t-tests analyzed silent pause rates between the utterances in the oral tasks and between the utterances in the picture task. There was no significant difference found for between-silent pauses between the oral and picture tasks, $t(41) = 1.105$, $p = 0.275$. The results suggest that between – silent pause rates were not inflated in the picture condition from the participants moving from one picture to the next. The locational comparison of silent pauses was exploratory and intended to answer a methodological concern. In the analysis described below silent pauses were analyzed as one maze type.

Maze Frequency

Table 1 displays the maze frequency (i.e., total number of mazes) for each age group in each elicitation task. More mazes were produced in the oral condition than the picture condition and more mazes were produced by the Kindergarten Group than the Grade Two group. Only one orphan was identified across groups and conditions. Normal distribution is not expected for the analysis of one orphan. Therefore, orphans were excluded from the analyses described below.

Table 1

Number of Mazes Produced by the Kindergarten and Grade Two Groups in Each Elicitation Task

Group and Task	FP	SP	RP	RV	O	Total
Kindergarten						
Oral	33	43	23	35	1	135
Picture	21	25	25	20	0	91
Total	54	68	48	55	1	226
Grade Two						
Oral	25	26	25	49	0	125
Picture	10	12	14	27	0	63
Total	35	38	39	76	0	188

Note. FP = filled pause; SP = silent pause; RP = repetition, RV = revision, O = orphan.

Overall Maze Rate

Age. The results indicate no significant main effect for age, F (1, 42) = 0. 272, p = 0.605, η_p^2 = 0.006. The Kindergarten and Grade Two groups produced mazes at similar rates. The mean maze rate for each group was 0.020 (SD = 0.023) and 0.018 (SD = 0.024) respectively.

Elicitation Task. The results indicate a significant main effect for elicitation task, F (1, 42) = 13.647, p = 0.001, η_p^2 = 0.245. More mazes were produced in the oral task than the picture task. See Table 2 for the mean maze rates in each task.

Table 2

Overall Mean and Standard Deviation of Maze Rates of the Kindergarten and Grade Two Groups in Each Elicitation Task

Elicitation Task	Group		
	Kindergarten	Grade Two	Total
Oral	0.024 (0.025)	0.021 (0.024)	0.023 (0.025)
Picture	0.016 (0.022)	0.015 (0.024)	0.015 (0.023)

Note. Mean is stated first, followed by standard deviation in brackets.

Age and Elicitation Task. The results indicate no significant interaction between age and elicitation task, F (1, 42) = 0.167, p = 0.685, η_p^2 = 0.004. The two groups produced mazes at similar rates in the picture task and in the oral task. See Table 2 for the mean maze rates for group and task.

Maze Type

Type. The results indicated a significant main effect for maze type, F (3, 40) = 3.872, p = 0.016, η_p^2 = 0.225. There were significant differences between the types of mazes produced by the participants across tasks. See Figure 1 for the mean maze rate of each maze type.

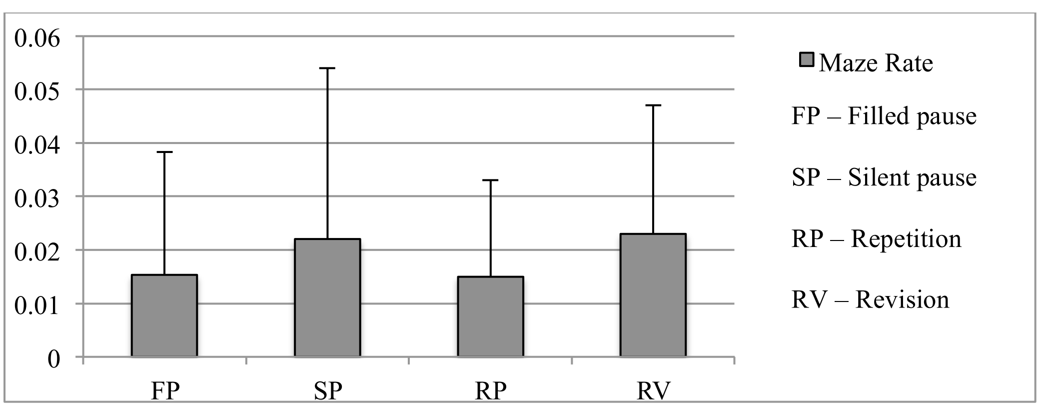

Figure 1: Mean maze rate and standard deviation for each maze type.

A series of paired samples t-tests were completed to establish the maze types that differed significantly across conditions and groups. A Bonferroni correction approach was used, with an adjusted p-value of 0.008. Alpha was divided by the number of tests performed (i.e. 0.05/6 = 0.008). The Grade Two participant with the unknown silent pause data was not included in the analysis of silent pauses for any of the t-tests. No maze types were found to differ significantly from one another. See Appendix C for the complete results from the comparison of maze types.

Age and Type. The results indicated no significant interaction between age and maze type, F (3, 40) = 1.641, p = 0.195, η_p^2 = 0.110. The Kindergarten and Grade Two groups produced maze types at similar rates. See Figure 2 for the mean maze rate of each maze type produced by each group.

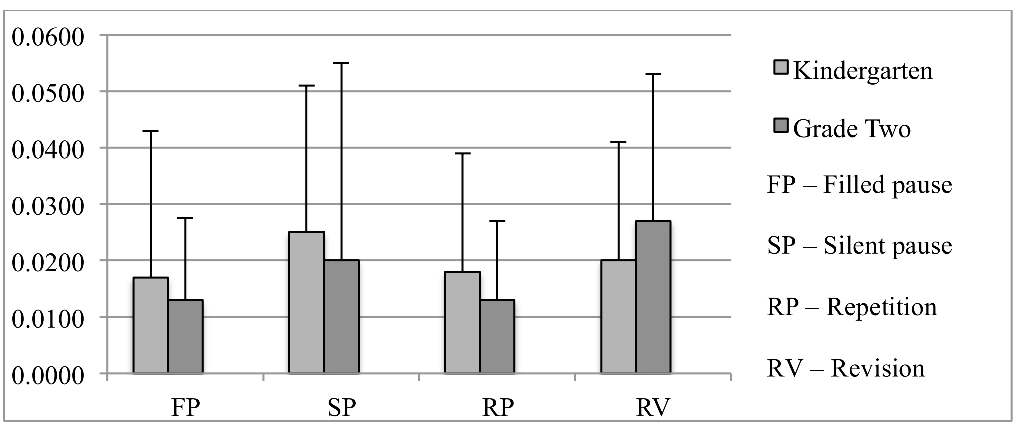

Figure 2: Maze rate and standard deviation for each maze type produced by Kindergarten and Grade Two participants.

Elicitation Task and Type. The results indicated no significant interaction between elicitation task and maze type, F (3, 40) = 1.441, p = 0.245, η_p^2 = 0.098. There was no difference between the types of mazes produced between tasks. See Figure 3 for the maze rate for each maze type produced in each elicitation task.

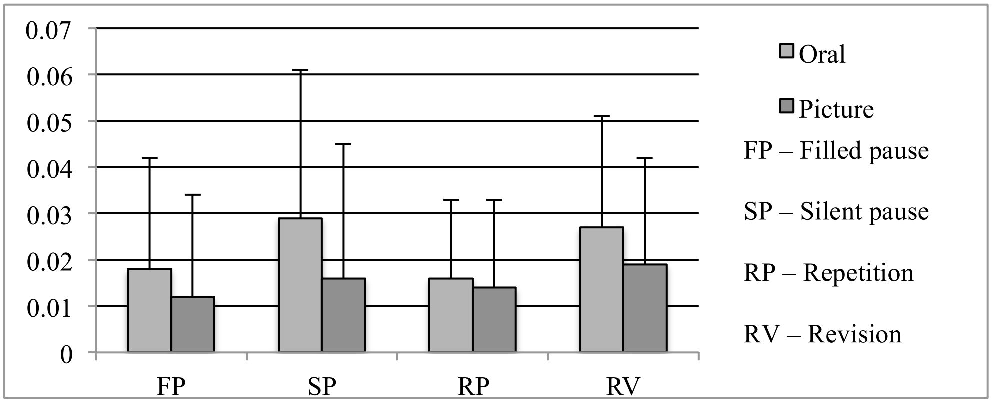

Figure 3: Maze rate and standard deviation produced for each maze type in the oral and picture elicitation tasks.

Age, Elicitation Task, and Type. The results indicated that there was no significant interaction between age, elicitation task, and maze type, F (3, 40) = 1.215, p = 0.317, η_p^2 = 0.084. The Kindergarten and Grade Two groups produced maze types similarly in each elicitation tasks. See Figure 4 for the maze rate for each maze type produced by each group in each elicitation task.

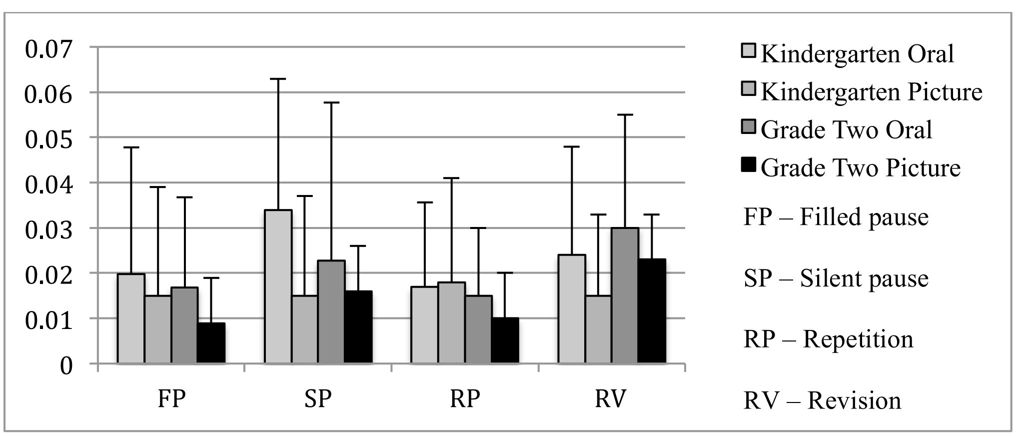

Figure 4: Mean maze rate and standard deviation for each maze type produced by the Kindergarten and Grade Two participants in the oral and picture elicitation tasks.

Discussion

The primary purpose of this study was to establish a reliable maze taxonomy based on speech behaviour that could be adopted by researchers with differing theoretical viewpoints. The maze taxonomy was developed and was found to be reliable. The second goal of this study was to add to the existing maze literature by using the maze taxonomy to examine age and task effects on maze production. The taxonomy was used to code the mazes in the stories of Kindergarten and Grade Two participants elicited from oral and picture tasks. The results indicated that the Kindergarten and Grade Two participants produced mazes at similar rates and used more mazes when telling stories elicited from oral tasks than picture tasks. No specific maze type differences were found between the tasks; however the directional difference between conditions for each maze type was higher in the stories elicited from the oral tasks than the picture tasks.

Maze Taxonomy

The reliability of the maze taxonomy used in this study was established. Two examiners coded mazes in 18 story transcripts and the agreement was high. Various studies have investigated the relationship between mazes and language ability, and mazes and language processes using a variety of methods to code and analyze mazes. Identifying maze patterns across these studies is difficult due to the significant differences in methodology. Several researchers noted the need to establish a unified method for coding mazes (Dollaghan & Campbell, 1992; Guo et al., 2008; Rispoli et al., 2008). Establishing a unified maze taxonomy is difficult because not all researchers have the same theoretical viewpoint.

Therefore, the logical direction appeared to be identifying a maze taxonomy based on speech behaviour rather than function or underlying theoretical factors. The labels used for the mazes in Dollaghan and Campbell's (1992) taxonomy describe the speech behaviour of the mazes rather than the underlying processes. This study used the Dollaghan and Campbell taxonomy as the framework for a new taxonomy. This new taxonomy adapted their definitions to be more transparent and succinct. The adaptations made to the taxonomy included separating filled and silent pauses into two main maze types, clarifying the repetition and orphan definitions, and omitting the coding and analysis of maze subtypes.

As will be described below, the maze taxonomy developed in this study is sensitive enough to identify maze production differences and could be used for coding mazes in other studies. A researcher could sort through the maze data from existing maze studies, code the mazes according to the taxonomy in this study, and compare the results. Perhaps other patterns or relationships will be identified. Whether or not it is adopted by future maze studies, it is a reliable and objective method for coding mazes in children's stories.

Task Effects on Maze Production

The participants in this study produced more mazes in oral tasks than in picture tasks. The results from this study support the findings of other studies (MacLachlan & Chapman, 1988; Schneider, 1996) that found a relationship between elicitation task and maze production. No specific maze types differences were found between the tasks; however, a directional trend was identified across all maze types. The maze rates for each specific maze type were higher in the stories elicited from the oral task than the picture task. The participants with typically developing language in this study and the participants with language impairments in Schneider (1996) produced mazes similarly in the oral and picture tasks. Participants in both studies produced more mazes in stories elicited from tasks with an oral component than stories elicited from a pictures-only task.

Story complexity and formulation difficulties may have contributed to maze production differences found between the tasks. In Schneider (1996) the oral elicited stories that the children with language impairments retold were better, more complex stories (i.e., contained more story grammars units) than the picture elicited stories. In Schneider and Dubé (2005), the original study from which the story transcripts in this study were obtained, the participants also told better, more complex stories (i.e., more story grammar units) in elicitation tasks with an oral component than the pictures only elicitation task. It appears that children, regardless of language ability, tell better, more complex stories when they have the opportunity to hear it first rather than when just presented with pictures. Children also produce more mazes in complex stories with an oral component than stories elicited just with a picture. This supports the notion that formulation load is higher for children when they hear a story than when they are presented with only pictures of a story (Schneider & Dubé, 2005).

Schneider (1996) hypothesized that children with language impairments encountered more formulation problems in the oral task than the picture task as evidenced by maze production. That hypothesis can be extended to the typically developing children in this study. As described above, the oral stories told by the participants were more complex than the stories elicited from the picture tasks. Story complexity is one contributing factor to the formulation difficulties the participants may have experienced in the oral task.

The formulation problems could also be related to several factors such as: story retell or construction demands; or memory load (Schneider, 1996). Retelling a story and constructing a story are two distinct tasks. The oral task was a story retell task and the picture task was a story construction task. It is possible that this difference contributed to the maze production differences. Memory load is possibly another contributing factor. During storytelling, pictures have been assumed to reduce the memory load (Schneider, 1996). When a storyteller is presented with pictures of stories she does not have to store as much information about the stories because she has the pictures to draw from. When a storyteller hears a story and is asked to retell it, she needs to store all the information of the story, resulting in a higher memory load.

A methodological concern was raised about the picture elicitation task used in this study. During the picture task it was possible that silent pauses between the utterances were inflated due to the time taken as the participants moved from picture to picture. If silent pause inflation occurred in the picture task then the silent pause rate between the utterances in the picture task would be higher than the silent pause rate between the utterances in the oral task. No significant differences were found between the two tasks for silent pauses that occurred between the utterances. Overall (although not statistically significant) more silent pauses were produced in the oral task than in the picture task. Given this discussion and the results of the data, silent pause inflation in the picture task is no longer a concern.

Age Effects on Maze production

No maze production differences were found between the Kindergarten and Grade Two participants in this study. The Kindergarten and Grade Two participants produced mazes at a similar rate and produced specific maze types at similar rates. This finding differed from Evans (1985) who found that Grade Two participants produced more mazes than the Kindergarten participants. In Evans the Grade Two participants also produced more revisions (i.e., corrections) and orphans than the Kindergarten students. Although Evans (1985) and this study compared the maze production between Kindergarten and Grade Two participants, and calculated maze rate similarly, the results were different. Methodological differences between the studies could account for the different results.

First, mazes were coded differently between studies. In Evans repetitions, revisions (i.e., corrections and postponements), and orphans (i.e., abandonments) were coded. Filled and silent pauses were not coded. In this study: filled pauses, silent pauses, repetitions, revisions, and orphans were coded. Another methodological difference between the studies was the type of elicitation tasks that were used. Evans used personal 'show and tell' type tasks to elicit language samples. In this study fictional stories were elicited from oral and picture tasks. Elicitation tasks influence maze production. It is likely that the maze production differences between the studies differed due to the elicitation tasks. The results of this study suggest that there is a need to further examine the relationship between task type and maze production at different ages.

Maze Types

Across participant groups and tasks there was a main effect for maze type. When the types were analyzed separately no differences were found. One explanation for this discrepancy is that there were too few participants to find statistical significance between the maze types. Although no significant maze type differences occurred, the maze production trend showed that revisions and silent pauses were produced at similar rates. Both were produced slightly more frequently than filled pauses and repetitions. Orphans had the lowest maze rate. Although there were no specific maze type differences that were found to be statistically significant the data suggest that there were production differences.

Orphans

One obvious maze type difference (although not statistically analyzed) was the orphan rate compared to the other maze types rates. Only one orphan was identified during this study- at the end of a story told by a Kindergarten participant, elicited from the oral task. The orphan definition established in this study and the nature of the task could account for this difference. The orphan definition used in this study was conservative. Orphans were defined as abandoned utterances that are followed by a new story/conversation. The new story/conversation could be preceded by a silent pause. The taxonomy states that if there is any doubt about whether the abandoned utterance is related to the original story/conversation it should be coded as revision rather than an orphan.

Other studies did not code orphans in this manner. For example, Evans (1985) coded orphans (i.e., abandonments) as the abandonment of an utterance and replacement with another utterance. According to the coding procedure in this study, the mazes Evans coded as orphans would likely be coded as revisions.

The nature of the tasks used in this study could also contribute to the low orphan rate. The storytelling tasks that were used were structured tasks. The participants either heard the story that they were asked to retell or they were provided time to construct the story that they would tell based on a set of pictures. The tasks did not require the participants to come up with a story on the spot as one would during a conversation or show-and-tell presentation. For example, in Evans (1985), the 'show and tell' task, where the children spontaneously spoke to their classmates about an item or experience, produced a high orphan rate. It is possible that children are more likely to abandon an utterance or word when they are presenting or talking about an item in a non-storytelling format. Whatever the case, it is still necessary to code for orphans in language samples, particularly because more orphans may be produced in other types of tasks (i.e., conversation or presentations).

Silent Pause Duration

The taxonomy used in this study coded and analyzed filled and silent pauses separately because they are two distinct maze types. Filled pauses contain verbal output while silent pauses do not. Although statistical significance was not found, more silent pauses were produced than filled pauses across participants and tasks. Even with the conservative duration method used in this study, silent pause rates had one of the highest rates across maze types. In the maze taxonomy established in this study, pauses without verbalization that were 2 seconds or longer were coded as silent pauses. The silent pause duration was adopted from Dollaghan and Campbell (1992) because it accounted for natural pause breaks that occur during conversation. The silent pause duration and method for calculating the duration in the maze taxonomy established by this study is conservative and reliable. It should be considered when establishing coding methods for silent pauses in future maze research.

Guo et al. (2008) found more silent pauses were produced than any other maze type across narrative tasks. Those results differ from the silent pause results of this study. One important difference between the studies was the silent pause duration that was used. Guo et al. used a fairly liberal method for coding silent pauses by including any break greater than 250 ms as a silent pause. The silent pause counts in Guo et al. most likely included natural pause breaks (e.g., pausing to take a breath or finish a sentence), pauses that are not considered to be mazes. Dollaghan and Campbell (1992) chose to include breaks that were 2 seconds or greater because most breaks are less than 2 seconds, many of which are thought to be natural pause breaks. It is likely that Guo et al.'s silent pause rates are inflated by natural pause breaks.

During the coding of the transcripts, silent pause durations were determined using the Audacity Software. Each silent pause was timed to the millisecond. This method was used because it appeared to be more objective and reliable than other methods (e.g., stopwatch). When inter-rater reliability was determined, there was disagreement around three silent pauses, one repetition, and one revision. Inter-rater agreement was still high across all maze types, but the disagreement of the three silent pauses indicates that even with an objective method, determining silent pause duration is difficult. Guo et al. (2008) used Multispeech software to determine silent pause duration. The procedure for determining duration with Multispeech software is similar to the procedure used with Audacity. It is likely that Guo et al.'s duration calculations are reliable. Other methods may not be as reliable, such as the stopwatch method used by Rispoli (2003). Methods for coding mazes need to be as objective as possible, including determining silent pause duration. The silent pause definition and method for determining silent pause duration used in this study are reliable and could be adopted for use by other maze studies interested in establishing reliable and objective silent pause rates.

Limitations

The main limitations of this study were the number of participants used, the close participant age groups, and the elicitation task design. It is likely that maze type differences do exist and were not identified statistically because there were not enough participants in this study. In the future, it is worth exploring whether maze types exist between age groups and between task types with larger participant groups.

Another limitation in this study was the close participant age groups. No maze production differences were found between age groups. It possible that there is a relationship between maze production and age, but the age groups compared in this study had similar language abilities so no maze production differences were found. More research is required to determine if there is a relationship between age and maze production. One way that this could be investigated further would be to use participants that differed more in age (e.g., maze rates of Preschool participants compared with the mazes rates of Grade Four participants). A longitudinal study would be the best indicator of whether a relationship between maze production and age exists.

To date, the elicitation tasks used in this study and in Schneider (1996) were the most controlled language elicitation tasks used among the maze literature. The story characteristics (i.e., story grammar units) were controlled in initial story presentation in each task. There are additional ways that elicitation tasks could be designed and controlled to target specific hypothesis related to maze production and underlying processes. For example, Guo et al. (2008) investigated the relationship between maze production and language ability by designing a study that matched participants for language ability and age. To investigate specific cognitive or linguistic processes, tasks should be designed to test those specific processes. For example, to investigate maze production and memory, controlled memory tasks could be used. To provide more information about the relationship between maze production and story complexity, oral stories controlled for story characteristics of varying complexity could be presented and then retold by the participants.

Future Directions

Mazes could provide insight into language ability and underlying language processes. To examine this, maze studies need to use controlled tasks that can be compared using objective maze coding procedures. The maze literature is currently small. A unified method for examining and coding mazes would provide a way of comparing results between studies. The suggestions for future maze research discussed below should be established on such a method. Comparing coding procedures between studies, analyzing speaker insight, examining listener perception, and studying mazes with other concurrent modes of expression could provide a more complete picture of language ability and underlying processes associated with mazes and communication.

In order to have a complete understanding of the effects that different maze coding procedures have on results, it would be beneficial to code the data from this study using coding procedures from other studies to determine if significant differences exist. The initial presentation of the storytelling task in this study controlled for story characteristics and the stories would provide a strong foundation for comparing coding procedures.

Another way to get at the underlying causes associated with maze production is through participant insight. In the future, analysis of the conversations between the participants and examiners should be studied. On some occasions participants had difficulty beginning or finishing a story and would tell the examiner that they could not remember the name of a character or some the story. Often times the participant would produce a filled pause, a silent pause, or a combination of pauses before or after admitting their forgetfulness (e.g., “And (:02.1) [SP] (um) [FP] (:04.9) [SP] I don't remember anything else”). The dialogue between the participant and examiner provides online insight into the language processes of the participants.

Listeners often know when someone is struggling to say something. Studying listener perceptions of the stories and storytellers could shed light into the common assumptions listeners and researchers make about speech fluency. Testing and analyzing listener perceptions and/or assumptions could point to underlying causes of maze production that have been overlooked by focused scientific methods.

Finally, examining maze production and other concurrent forms of expression together (e.g., gestures and facial expression) could provide a more complete picture into the underlying processes associated with speaking or storytelling. Many theorists and maze researchers subscribe to the theory of cognitive load. They suspect that mazes are external representations of glitches or a taxing of the language and/or cognitive system. In gesture research, the cognitive load theory has been used (Goldin-Meadow, 2007) to explain gesture use. When a speaker is talking, gestures are thought to reduce the cognitive load, freeing up cognitive effort that can be used for other tasks. Like mazes, gestures may be another external indicator of the language and/or cognitive systems that are being taxed. It is possible that mazes and gestures provide a speaker with some cognitive or processing relief when they are telling stories or conversing.

Conclusion

Few relationships have been identified between maze production and language ability, and maze production and language processes. This study addressed the methodological inconsistencies within maze literature responsible for conflicting results. The primary purpose of this study was to develop a reliable maze taxonomy based on speech behaviour rather than underlying processes. This approach would enable researchers with different theoretical viewpoints to use the taxonomy. The maze taxonomy was constructed and was found to be reliable. The second goal of the study was to add to the current maze literature by investigating age and task effects on maze production. The results did not indicate age effects, but task effects were found. More mazes were produced in stories elicited from oral tasks than picture tasks. Story complexity and formulation difficulties could have contributed to the maze production differences found between the tasks. More research is required to confirm this hypothesis. No specific maze type differences were found, but that is likely due to the small number of participants used in this study.

The maze taxonomy established in this study answered the call from other maze researchers (Dollaghan & Campbell, 1992; Guo et al., 2008; Rispoli et al., 2008) that requested a unified method for coding mazes. A coding procedure was developed that could be adopted by researchers with differing theoretical viewpoints. In conjunction with the maze taxonomy developed, maze researchers need to design studies that will test specific hypotheses related to mazes and language ability or processes. If other maze researchers adopt this taxonomy, and create strong language elicitation tasks, it is likely that consistent maze patterns will emerge.

References

Audacity (Version 2.0.3) [Computer software]. Retrieved March 17, 2013, from http://audacity.sourceforge.net

Bortfeld, H., Leon, S., Bloom, J., Schober, M., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*, 123-147.

Dollaghan, C., & Campbell, T. (1992). A procedure for classifying disruptions in spontaneous language samples. *Topics in Language Disorders*, *12*, 56-68.

Evans, A. (1985). Self-initiated speech repairs: A reflection of communicative monitoring in young children. *Developmental Psychology*, *21*, 365-371.

Guo, L., Tomblin, J.B., & Samelson, V. (2008). Speech Disruptions in the narratives of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *51*, 722-738.

Kowal, S., O'Connell, D. C., & Sabin, E. J. (1975). Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, *4*, 195-207.

Mayer, M. (1977). *Oops*. New York: Dial Books for Young Readers.

Miller, J. F. (2002). Systematic Analysis of Transcripts (SALT, Version 7) [computer software]. Madison, University of Wisconsin, Waisman Research Center.

Navarro-Ruiz, M., & Rallo-Fabra, L. (2001). Characteristics of mazes produced by SLI children. *Clinical Linguistics and Phonetics*, *15*, 63-66.

Nettelbladt, U., & Hansson, K. (1999). Mazes in Swedish pre-school children with specific language impairment. *Clinical Linguistics and Phonetics*, *13*, 483-497.

Olmos, A. (2007) Cohen's kappa index: Inter and intra reliability of a method. In *Diagnostics in phytopathology*. Retrieved March 28, 2013, from http://antonio-olmos.com/kappa/concepts.html.

Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. *Journal of Speech, Language, and Hearing Research*, *46*, 818-830.

Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech, Language, and Hearing Research*, *51*, 953-966.

Schneider, P. (1996). Effects of pictures versus orally presented stories on story retellings by children with language impairment. *American Journal of Speech-Language Pathology*, *5*, 86-96.

Schneider, P., & Dubé, R. V. (1997). Effect of pictorial versus oral story presentation on children's use of referring expressions in retell. *First Language, 5 (3),* 283-302.

Schneider, P., & Dubé, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology*, *14*, 52-60.

Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

Stein, N. L., & Glenn. C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing*, *Vol. 2: Advances in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.

Thordardottir, E., & Weismer, S. (2002). Content mazes and filled pauses in narrative language samples of children with specific language impairment. *Brain and Cognition*, *48*, 587-592.

Wagner, C., Nettlebladt, U., Sahlen, B., & Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders*, *35,* 83-93.

Zackheim, C. T., & Conture, E. G. (2003). Childhood stuttering and speech disfluencies in relation to children's mean length of utterance: a preliminary study. *Journal of Fluency Disorders*, *28*, 115-142.

Appendix A

Maze Coding Procedure Adapted From Dollaghan and Campbell (1992)

Filled Pauses: Any non-lexical, one syllable fillers that occur between words and/or mazes [FP].

- Coding instructions: Code any filled pauses that occur between and within utterances (e.g., (e.g., "(um) [FP] Once a upon a time (um) [FP] there was a hippo"). Code multiple successive filled pauses as one occurrence by placing all of the filled pauses in one parentheses and label them with one code (e.g., "One day (ah ah) [FP] she was walking down the tracks"). If a filled pause occurs during a revision, include both mazes within the same parentheses, and code both types of mazes (e.g., "(And then um she get) [FP] [RV] Then she saw a nice flower in the train track").

Silent Pauses: A silent break that is 2 seconds or greater in duration [SP].

- Coding instructions: Code any silent pauses that occur between and within utterances (e.g., "(Um) [FP] (:02) [SP] the (:02) [SP] oranges came crashing down"). Determine the duration of each silent break with Audacity Software and if the pause 2 seconds or greater, mark the silent pause with a colon followed by the length of the pause (e.g., "(:02) [SP] One day the hippo went to the grocery store"). If a silent pause occurs during a revision, include both mazes within the same parentheses, and code both types of mazes (e.g., "(So then :03 the hippo) [RV] [SP] so then Heather the hippo went to the store").

Repetitions: The exact repetition of a linguistic unit of any length [RP].

- Coding instructions: Code multiple successive repetitions as one occurrence by placing all of the repetitions in one parentheses and labeling them with one code (e.g., “Then (m*, Mr.) [RP] Mr. Penguin got angry”). Code filled pauses and silent pauses between the original utterance and the repeated portion of the utterance, code accordingly (e.g., “Heather the hippo (took an or*) [RP] (um) [FP] orange from the stand”). If a repetition occurs during a revised maze code both mazes within the same parentheses and mark both types of mazes (e.g., “(Heather the hungry hi hi hippo) [RP] [RV] heather the hippo went to the museum”). Do not code the maze as a repetition if any linguistic information is added or deleted to the beginning or middle portion of the repeated utterance (e.g., “(He's) [RP] (He's) [RV] He was standing by the oranges”). Do not include words that are repeated for emphasis as repetitions (e.g., “Heather the Hippo really, really liked the flower”). Words that are repeated for emphasis are normally describing words and often have a change in intonation and/or an increase in volume. Often repetitions that are mazes have the same intonation and volume.

Revisions: Recognizable modifications of a linguistic unit already produced [RV].

- Coding instructions: Code any additions or omissions of sounds, words, phrases, or whole utterances to a linguistic unit already produced as revisions (e.g., “(He’s) [RV] and he’s standing in the store”). When there are multiple successive revisions, code them in separate parenthesis and label them with separate codes (e.g., "The bones (w*) [RV] (it) [RV] began to fall"). Filled pauses, silent pauses, and repetitions can occur between the original utterance and the revised portion of the utterance (e.g., "(And then she) [RV] (um) [FP] (and then the elephant) [RV] and then Mr. Elephant got mad." or "All of the sudden (the oranges started) [RV] (the) [RP] the oranges began to fall").

Orphans: Abandoned utterances that are followed by a silent pause or a new story/conversation that does not return to the original story or utterance [O].

- Coding instructions: Code abandoned utterances that are followed by a silent pause or a new story/conversation (e.g., "(And then) [O] (um um) [FP]… I think I am ready to go back to class"). If there is linguistic information that follows an abandoned utterance and it is not a new story or conversation, do not code it as an orphan, code it as a revision. If there is any doubt that the original utterance is or is not related to the linguistic information that follows it, code the maze as a revision (e.g., "(Sh) [RV] (:05) and the store"). Do not label a maze as an orphan if the utterance is followed by a comment to the experimenter and the child returns to telling the story after the comment. In that case, code the maze according to the other maze descriptions (e.g., "(:03) [SP] (She was a) [RP] {To A1: I forgot}. She was a hippo").

Appendix B

Inter-rater Agreement for Coding Mazes By Type

		FP	WSP	BSP	RP	RV	O	Non-Maze	Total
Rater B	FP	22							15
	WSP		8					1	9
	BSP			18					18
	RP				15				15
	RV					33			33
	O								0
	Non-Maze			2	1	1			4
	Total	15	8	20	16	34	0	1	101

Note. FP = filled pause; WSP = within-silent pause; BSP = between-silent pause; RP = repetitions; RV = revisions; O = orphans.

Appendix C

T-Test Results Comparing the Mean Maze Rate of Maze Types

Pair	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)	Difference (p < 0.008)
FP – SP	-0.007	0.038	0.004	-1.646	85	.103	No
FP – RP	-0.000	0.025	0.003	-0.025	87	0.98	No
FP – RV	-0.008	0.027	0.003	-2.572	87	0.012	No
SP – RP	0.007	0.034	0.037	1.748	85	0.084	No
SP – RV	-0.001	0.041	0.005	-0.300	85	0.765	No
RP – RV	-0.008	0.028	0.030	-2.645	87	0.010	No

Note. Std. Deviation = standard deviation; Std. Error Mean = standard error mean; df = degrees of freedom; FP = filled pause; SP = silent pause; RP = repetition; RV = revision.