The Influence of Gesture Production upon Reading Comprehension Outcomes across Development

by

Lisa Jean Smithson

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology
University of Alberta

Abstract

When engaged in a reading comprehension task, the reliance upon a verbal or visuospatial strategy is associated with different outcomes (Johnson-Glenberg, 2000). More specifically, individuals who adopt a verbal strategy tend to demonstrate enhanced recall of information that is explicitly stated in the text while individuals who adopt a visuospatial strategy tend to demonstrate enhanced recall of information that must be inferred from a text (Johnson-Glenberg, 2000). Previous research has shown that gesture production can influence strategy use in a problem-solving task (Alibali, Spencer, Knox, & Kita, 2011). The purpose of the current research was to investigate whether gesture production influences reliance upon a verbal or visuospatial strategy in a reading comprehension task and to determine whether gesture production influences reading comprehension outcomes. Previous research has demonstrated that the generation of text summaries improves reading comprehension (McKeown, Beck, & Blake, 2009). In Study 1, adult participants were presented with easy and standard reading passages that were divided into three paragraphs. Participants were randomly assigned to communicate everything that they could remember from each of the paragraphs in one of four experimental conditions: (1) *Gesture*: participants were encouraged to use meaningful hand gestures during summary generation, (2) *Restricted*: participants were restricted from moving their hands during summary generation, (3) *Control:* participants were not provided with any instructions regarding their movement during summary generation, (4) *Written*: participants were asked to write down their summaries. Measures of vocabulary, verbal working memory, visuospatial working memory, and motivation were obtained. Among participants in the Gesture and Control conditions, visuospatial working memory and motivation measures were the best predictors of reading comprehension outcomes, whereas among the participants in the Restricted condition, vocabulary was the best predictor of reading comprehension outcomes. Though individuals in the Gesture condition did not experience any notable reading comprehension outcome advantages, individuals who gestured spontaneously in the Control condition did. Further analyses revealed that verbal working memory was a negative predictor of gesture rate. It was hypothesized that gesture production may only be beneficial in the context of a reading comprehension task when verbal working memory resources are taxed. In Study 2, this hypothesis was tested among children by using reading passages of three levels of increasing difficulty. Children were randomly assigned to a *Gesture* or *Control* condition. Children in the Gesture condition demonstrated an advantage on both measurements of reading comprehension that did not require the generation of inferences and those that did. In conclusion, these findings suggest that gesture production can influence strategy use in a reading comprehension task. Research methodologies used to assess reading comprehension vary with respect to whether participants are free to move (e.g., Chinn, Anderson, & Waggoner, 2001) or are restricted from moving (e.g., Humphreys & Gennari, 2014). The results from this study suggest that these methodologies may promote different strategies, thereby biasing research in this field. The results from this study also suggest that gesture production may be a useful strategy for individuals who struggle with verbal skills. When children have difficulty understanding text, it may be useful for caregivers and teachers to encourage them to use their hands to represent the ideas presented in the text in meaningful ways.

PREFACE

This thesis is an original work by Lisa Smithson. The research projects, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board. 'Visuospatial and verbal working memory, body positioning, and reading comprehension among adults', 36020, January 4th/2013 and 'Visuospatial and verbal working memory, body positioning, and reading comprehension among children', 36047, December 18th/2012.

ACKNOWLEDGMENTS

There are many people who contributed generously to this work. I would like to thank all of the adults and children who participated in this research. My interactions with these participants brought this research to life for me. These interactions have also inspired more research questions than I could possibly investigate in my lifetime. I would also like to thank the teachers and the principals of the children who participated in this research, for welcoming my research assistants and I into their schools. These teachers and principals enthusiastically went out of their way to accommodate this work, and their support is appreciated.

I would like to thank my supervisor, Elena Nicoladis, for challenging me to grow as a researcher and for playing such a pivotal role in making my graduate studies such a fun and fulfilling experience. I would also like to thank Sandra Wiebe and Jeremy Caplan for providing me with encouragement and helpful feedback along the way.

Jasmine Aziz, Suzette Duong, Jessica Lee, and Max Millions all invested countless hours into running research sessions and coding data. The generous help from these research assistants contributed significantly to this work.

I would like to thank my parents, Elizabeth and Sidney Smithson, for being such dedicated and enthusiastic cheerleaders throughout this journey. Their support played a critical role in bringing this dissertation into existence.

I would like to thank James Farley for keeping me relatively sane throughout the writing process. I would also like to thank Anna Bencetic, Jennifer Carroll, Mike Flannigan, Jennifer Passey, Carley Piatt, May Warren, and Sandra Ziolkowski for all of their encouragement and advice, especially throughout the final leg of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction and Literature Review

The goal of reading comprehension is to "gain an overall understanding of what is described in the text rather than to obtain meaning from isolated words or sentences" (Woolley, 2011, p.15). The purpose of the current research is to investigate how gesture production while summarizing information that has been read, influences reading comprehension outcomes among adults and children.

Broader Context of the Current Research

When understanding written texts, individuals develop mental models of the text ideas (Kintsch, 1998). These are representations of the meaning in the text. Two classes of mental models are thought to exist which highlight the dichotomy between verbal and visuospatial reading comprehension strategies: (1) a text-based model which consists of propositions contained within the text without adding any information that is not mentioned explicitly in the text and (2) a situation model which consists of understanding the information presented within the text by making inferences about the meaning of the text using the individuals' knowledge and experience (Kintsch, 1998). The text-based model incorporates propositions from the text with inferences that are required to bring coherence to the text, and retains verbatim information that is contained within the text (Kintsch, 1998). In contrast, situation models consist of elaborative inferences and link personal knowledge with the text information (Kintsch, 1998). Situation models do not typically retain verbatim information contained within the text but instead lend support to a more flexible structure of knowledge (Kintsch, 1998). According to Woolley (2011), the primary difference between a text-based and situation model is that the text-based model depends upon the generation of few inferences, whereas the situation model depends upon the generation of many inferences. Reliance upon text-based and situation models may differ depending upon whether an individual adopts a verbal or visuospatial strategy during reading comprehension.

Verbal and Visuospatial Strategy Adoption in Cognitive Tasks

People use a variety of strategies in order to complete cognitive tasks. These strategies can differ in terms of their reliance upon verbal and visuospatial modes of information processing. For example, in order to remember a phone number, you may repeat the individual numbers subvocally, or you may mentally visualize the sequence of numbers. Verbal and visuospatial strategy preference has been found to differ across development (Palmer, 2000) and as a function of the particular problems being solved (Trbovich & LeFevre, 2003). Paivio's (1986) Dual Coding Theory has been important in providing an understanding of how verbal and nonverbal systems can be used during reading comprehension. According to the *Dual Coding Theory*, cognition is comprised of activity that occurs in two mental codes: a verbal code that is reliant upon language and a non-verbal code that is reliant upon mental imagery for the coding of both objects and events (Sadoski, Paivio, & Goetz, 1991). From this point forward these codes will be referred to as verbal and visuospatial codes, respectively. In memory, these two types of mental codes are represented differently. Verbal information tends to be encoded in a sequential manner while visuospatial information tends to be encoded in a holistic manner (Sadoski et al., 1991). Reliance upon verbal and visuospatial representations tends to lead to different strengths in information processing. Verbal representations tend to be superior with respect to the representation of abstract knowledge and the organization of thought (Kosslyn 1988; Kosslyn et al., 1995a). Visuospatial processing tends to be more concrete and encompassing (Kosslyn 1988; Kosslyn, Behrmann, & Jeannerod, 1995a). Begg and Anderson (1972) argue that abstract ideas tend to be more likely stored in verbal representations, while concrete phrases tend to be remembered imagistically. Researchers have highlighted the dichotomy between verbal and visuospatial strategies in the context of reading comprehension (Kirby, 1993; Mayer & Sims, 1994; Plass, Chun, Mayer, & Leutner, 1998; Shah & Miyake, 1996). Verbally encoded and visuospatially encoded information both play critical and somewhat unique roles in reading comprehension. What factors influence reliance upon verbal and visuospatial codes in a reading comprehension task? To address this question it is first necessary to understand how verbal and visuospatial processing contribute to reading comprehension.

Contributions of the verbal code to reading comprehension. A variety of verbal skills including phonological awareness, phonemic knowledge, orthographic knowledge, and alphabetic reading skill are critical for effective reading comprehension (Woolley, 2011). Importantly, it is not just the breadth of word knowledge, but rather the depth and quality of lexical representations in memory that predict reading comprehension skills (Woolley, 2011). Word decoding and vocabulary are important for comprehension, but the ability to process words in longer discourse is required. Individuals with delays in reading often require instruction with both vocabulary development and word decoding (Bishop, 1997; Stanovich, 1986). Vocabulary, word decoding, and verbal working memory will be described here in terms of their contributions to reading comprehension. It is important to note that vocabulary, verbal working memory and word decoding measures have been found to make significant independent contributions to reading comprehension when assessed collectively as predictors (Seigneuric, Ehrlich, & Oakhill, 2000). In previous research, vocabulary has been found to be the best predictor followed by verbal working memory and then decoding speed (Seigneuric et al., 2000).

Vocabulary. Text comprehension is strongly reliant upon knowledge and familiarity of word meanings. Encountering unfamiliar words when reading can hinder comprehension and the coherence of the discourse (Perfetti, 1994). Individuals with weak vocabularies tend to find reading difficult and avoid reading (Juel, 1998). Because of this avoidance, they encounter fewer new words and develop more limited vocabularies (Juel, 1998). Skilled readers are able to derive meanings for unknown words by using words that appear in the context. Acquiring these new meanings leads to vocabulary growth (Neal & Kelly, 2002; Swanborn & de Glopper, 2002; Worthy, Patterson, Salas, Prater, & Turner, 2002). There is a strong and positive association between vocabulary and reading comprehension (Cain, 2007; Gersten, Fuchs, Williams, & Baker, 2001; Oakhill & Cain, 2007; Pearson, Hiebert, & Kamil, 2007). Furthermore, vocabulary development is one of the strongest predictors of reading comprehension (Bowers & Sunseth, 2002).

Verbal working memory. Measures of verbal working memory tend to be positively correlated with the temporary storage and manipulation of sound and language information (Baddeley, 2003). According to Baddeley (1990), verbal working memory tends to store information in a linear and sequential manner: like speech, the verbal information is not represented simultaneously but rather as a sequence of information. Researchers have found that verbal working memory measures tend to positively predict reading comprehension among children (Cain, 2006; Cain, Oakhill, & Bryant, 2004; Daneman & Marikle, 1996; Goff et al., 2005; Just & Carpenter, 1992; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Oakhill, Cain, & Bryant, 2003; Seigneuric et al., 2000; Swanson & Howell, 2001; c.f. Dufva, Niemi, & Voeten, 2001; Oakhill, Yuill, & Parkin, 1986) and among adults (Daneman & Carpenter, 1980; Jincho, Namiki, & Mazuka, 2008; St. Clair-Thompson, 2007). Verbal working memory may be particularly useful for reading comprehension since it may enable the reader to use context more effectively in order to understand words within a text (Daneman & Green, 1986). For example, when encountering new words, readers with strong scores on verbal working memory assessments remember contextual information provided in the text more effectively than individuals with weak scores on verbal working memory measures. This activation of contextual information can help an individual to decipher the meaning of unfamiliar words.

Decoding speed. Decoding speed is one index of fluency (Woolley, 2011). Word-reading fluency is the ability to quickly and correctly recognize words in a written form (Perfetti, 2007; Stanovich, 1986). According to Nichols, Rupley, and Rasinnski (2009), fluency allows for the integration of information in the written context since it helps readers to process large strings of connected discourse. More specifically it is thought that the improvement of fluid word-reading shifts readers' attention from sub-lexical units towards higher language and cognitive processes which allow for the comprehension of large units of text (Bashir & Hook, 2009; LaBerg & Samuels, 1974; Perfetti, 2007). An increase in decoding speed is associated with improved reading comprehension (Klauda & Guthrie, 2008).

Contributions of the visuospatial code to reading comprehension. When reading, people internally simulate the events that they read about (Barsalou, 1999; Bower & Morrow, 1990; Zwaan, 2004; Zwaan & Radvansky, 1998). A variety of visualization strategies have been used in order to enhance reading comprehension. Illustrations (Kendeou, Savage, & Van den Broek, 2009), drawing (Kintsch, 1994) and the meaningful manipulation of physical objects that are relevant to text material (Glenberg, Brown, & Levin, 2007) have been useful in improving reading comprehension. Mental imagery can also be effective in improving reading comprehension (Sadoski & Quast, 1990). Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004) investigated the usefulness of an intervention technique designed to assist children in creating mental simulations. In this experiment, imagined manipulation of toys in meaningful ways enhanced memory for the text. According to Glenberg et al. (2004), imagined manipulation can be useful since it assists children in gaining an understanding that goes beyond the information provided explicitly in the text by incorporating inferences and real-world knowledge.

The Indexical Hypothesis (Glenberg & Robertson, 1999; 2000) may clarify why simulation has a beneficial effect upon reading comprehension. According to the Indexical Hypothesis, “language is made meaningful by cognitively simulating the actions implied by sentences” (Glenberg & Kaschak, 2002, p. 559). Three processes are required in order to transform words and syntax into meaning that is action-based: (1) the words are indexed (i.e., the abstract individual words are indexed to a perceptual symbol), (2) affordances are derived for the perceptual symbols that were generated during the indexing of the words, and (3) the affordances are integrated in a way that is guided by sentence syntax. If the affordances are not integrated in such a way that they could guide action, then understanding of the text is incomplete. According to Glenberg et al. (2004), actual and imagined manipulation can facilitate the derivation of affordances that are necessary to build an integrated understanding of what has been read.

Van Meter et al. (2006) commented similarly upon the usefulness of the linkage of verbal and visuospatial representations in working memory. According to Van Meter, Aleksic, Schwartz, and Garner (2006), “it is the verbal representation that serves as the foundation for the construction of the nonverbal representation” (p. 145). The integration of the verbal and visuospatial representations allows for superior performance on assessments of reading comprehension that require knowledge application, but does not tend to facilitate performance on reading comprehension tasks that require verbatim recognition (Van Meter et al., 2006).

Visuospatial working memory. Measures of visuospatial working memory tend to be positively correlated with the temporary storage and manipulation of visual and spatial information (Baddeley, 2003). Researchers argue that this memory system is responsible for imagery generation (Cattaneo, Fastame, Vecchi, & Cornoldi, 2006) and tends to store visual and spatial information holistically and concurrently (Alloway, Gathercole, & Adams, 2004). Some researchers have found that measures of visuospatial working memory positively predict reading comprehension (e.g., Goff et al., 2005; St. Clair-Thompson, 2007) while others have not (e.g., Seigneuric et al., 2000; Swanson & Howell, 2001). One possible reason for these mixed findings is that a variety of assessments are used for reading comprehension: some that require the recall of verbatim information and others that do not. For example, measures of visuospatial working memory tend to positively predict reading comprehension when multiple-choice questions are used (e.g., Goff et al., 2005; St. Clair-Thompson, 2007) but not when fill-in-the-blank questions are used (e.g., Seigneuric et al., 2000; Swanson & Howell, 2001).

Comparing the effectiveness of verbal and visuospatial reading comprehension interventions. Johnson-Glenberg (2000) explored whether verbal or visuospatial strategy training interventions are more beneficial for poor reading comprehenders. In order to train the verbal code, a training package of four verbal strategies was used: summarization, clarification, prediction, and question generation. In order to train the visuospatial code of processing, students were trained to "make movies in their heads" (p.773) and to discuss the images that they had created and discuss text summaries with a group. The verbal training program group performed significantly better than the control group on a test of word recognition, question generation, and answering explicit open-ended questions (open-ended in the sense that no answer options were provided), but did not differ from the control group on a measure of implicit open-ended questions (though the effect was marginal in favor of the verbal training group). The explicit open-ended questions required students to provide answers to questions that referred to information that was explicitly stated in the text. The implicit open-ended questions required students to make inferences in order to arrive at the correct answer. The visuospatial training program group performed significantly better than the control group on answering implicit open-ended questions. The verbal group significantly outperformed the visuospatial group on answering open-ended explicit questions. They also displayed marginally better performance on the question generation and listening recall tasks. The findings from the study by Johnson-Glenberg (2000) suggest that the verbal group was more efficient than the visuospatial group at recalling information that was factual, linear, and explicitly stated in the text. Importantly, although both groups answered significantly more implicit open-ended questions than the control group, this difference was larger for the visuospatial group. In conclusion the authors note that both verbal and visuospatial strategies tend to enhance word recognition and performance on open-ended questions. They also note that the question of whether verbal or visuospatial strategies are superior for improving reading comprehension performance cannot be answered clearly with these data. Instead, they argue that recalling information from a text is an important component of reading comprehension and tends to be more strongly enhanced with a verbal training program. In contrast, creating inferences tends to be the foundation of advanced reading comprehension abilities, and this skill tends to be more strongly enhanced in a visuospatial training program.

Meaningful Movements may Affect Reliance upon the Visuospatial Code

What factors influence whether an individual will adopt a reading comprehension strategy that is more reliant upon a verbal or visuospatial code? Movement of the body will be explored here as a possible predictor of the reliance upon a visuospatial code. The thesis that will be tested is that movements of the body can influence the degree to which a verbal or visuospatial code is adopted in reading comprehension tasks among adults. This thesis is based upon the theory of embodied cognition. Barsalou (1999) has proposed the Perceptual Symbol Systems theory in an effort to provide a perceptual theory of knowledge grounded in embodied cognition. Barsalou argues that when people are engaged in perceptual experiences, bottom-up patterns of activation in sensory-motor areas are captured by the association areas of the brain. When later recalling this experience, people can use top-down patterns of activation to partially reactivate these sensory-motor areas of the brain. This partial reactivation forms the basis of simulations. Individuals can demonstrate creativity by simulating experiences that they have never experienced before by combining perceptual symbols in novel ways. These simulations can even defy properties that exist in the physical world. Evidence for this theory has accumulated. Visual imagery is often accompanied by activation in the primary visual cortex and other early visual areas (e.g., Kosslyn, Thompson, Kim, & Alpert, 1995), auditory imagery is often accompanied by activation in early auditory areas (e.g., Zatorre, Halpern, Perry, Meyer, & Evans, 1996), and motor imagery is often accompanied by activation in the primary motor cortex and other early motor areas (e.g., Crammond, 1997; Jeannerod 1994; 1995). Motor imagery is also associated with limb movements (see Guillot & Collet, 2005 for a review) and heart rate (Decety, Jeannerod, Durozard, & Baverel, 1993). Furthermore, damage to a sensory-motor area leads to category specific deficits in the processing of conceptual information (e.g., Gainotti, Silveri, Daniele, & Giustolisi, 1995; Pulvermüller, 1999). The usefulness of embodied thinking in language processes has been demonstrated with the action-sentence compatibility effect (ACE) (Glenberg & Kaschak, 2002). When participants assume body movements that coincide with the meaning of a sentence, they are faster to make sensibility judgments. For example, if people are presented with the sentence “close the cupboard”, participants will respond more quickly when sensibility judgment responses are made by moving the hand away from the body instead of moving the hand towards the body. This research suggests that meaningful movements of the body may facilitate reading comprehension by capitalizing upon simulation generation.

Gesture production and imagery. One common type of meaningful body movement is gesture production. Gestures are hand and arm movements that people produce when they are engaged in effortful cognitive activities (Alibali, 2005). Gestures are one form of movement that may be strongly associated with information processing in an imagistic code. Hostetter and

Alibali (2008) have proposed the Gesture as Simulated Action model in order to explain how gestures are produced according to an embodied framework. According to this model, the simulation of actions involves the activation of premotor states and this activation can spread to motor areas and become externalized in an overt action. This spreading activation is thought to be responsible for the production of gestures. Four lines of research have emerged to support a strong association between gesture production and imagery: (1) gesture production tends to be associated with the level of imagistic content in the words used in speech, (2) gesture production tends to increase when the demands on imagistic resources increase, (3) gesture production tends to enhance the recall of imagistic information, and (4) gesture production is associated with the adoption of imagistically based strategies in problem solving tasks.

Gesture and imagery in speech. Research suggests that gesture production is associated with imagistically rich speech. For example, in a study by Rimé, Shiaratura, Hupet, and Ghysselinckx (1984), participants were asked to engage in a 50-minute conversation with an experimenter. In this experiment, participants conversed with the experimenter while being restricted from moving and while being free to move. More specifically, three phases were involved: during the first 15 minutes, the participant was free to move, for the next 20 minutes participants were restricted from moving and for the last 15 minutes participants were free to move again. In the movement restricted phase of this experiment, participants' hands, arms, head, legs, and feet were restricted. Participant speech was assessed by a computer program, which quantified the degree of speech imagery used in each of the phases. The results of this study indicated that when body movements were restricted, the imagery index of the speech was significantly lower than when movement was permitted. The researchers argued that a speaker's nonverbal behavior represents an important component of ongoing representational processes. Movements of the body may activate imagistically rich thoughts (Rimé et al., 1984).

The study by Rimé et al. (1984) suggests that gesture production may lead to more imagistic speech. However, since much of the body was restricted from moving it is difficult to know whether movements specific to the hands and arms as opposed to movements of the body more generally, are associated with imagistic speech. Research by Rauscher, Krauss, and Chen (1996) lends support to the argument that hand and arm movements are associated with imagery in speech. In this study the effects of gesture restriction were assessed in relation to the presence of visuospatial information in speech. If gesture production preferentially activates imagistic speech, then restricting gestures should hamper imagistically rich speech more than speech that is not imagistically rich. Six videotaped excerpts (averaging 3 minutes each) from a cartoon were used as the stimuli for this study. Participants described 3 video excerpts with electrodes placed on their palms (the gesture restricted condition) and they described 3 video excerpts with electrodes placed on their ankles (the gesture condition). Participants were told that these electrodes were measuring electrodermal activity and were instructed to keep their hands on the electrodes (which were fixed to the armrests of their chair) at all times in the gesture restricted condition (note that electrodermal activity was not actually recorded in this study). Participant narratives were assessed for phrases that contained visuospatial prepositions. The results indicated that participants spent significantly more time gesturing during phrases that contained visuospatial prepositions in comparison to other phrases. Speech rate was reduced when speakers were referring to visuospatial content and were prevented from gesturing. When the content was not visuospatial, the speakers spoke more quickly when they were restricted from gesturing. Additionally, when the content of speech was visuospatial, gesture restriction increased the rate of dysfluencies but when the content of speech was nonspatial, gesture restriction did not affect the rate of dysfluencies. The researchers suggest that the results of this study highlight the interaction that exists between gesture and speech production systems. Taken together, the studies by Rimé et al. (1984) and Rauscher et al. (1996) suggest that gesture production affects the imagery content of speech.

Gesture production increases when the demands on imagistic resources increase. If gesture production activates imagistic resources preferentially, then tasks that place a stronger demand upon imagistic resources should result in a greater reliance upon gesture production. This has been illustrated in a variety of studies. In a study by Morsella and Krauss (2004), participants described objects that were either present or absent. A mixed 2x2x2 design was used in which participants were either assigned to a gesture allowed or restricted condition (between subjects factor) and were either assigned to describe objects that were present or absent from view (between subjects factor). Additionally, all participants described both codable (i.e., objects that can be labeled easily, such as a star) and noncodable stimuli (i.e., nonsense figures that could not be labeled easily, such as a cluster of squiggles) (within subjects factor). Four important results emerged from this study. Firstly, gesture rate was significantly higher when the object was absent from view in comparison to when it was present. This finding served as a replication of previous research (De Ruiter, 1998; Wesp, Hesse, Keutmann, & Wheaton, 2001) and suggests that gesture production may have facilitated the temporary storage of visuospatial information when the objects were absent. Secondly, individuals still gestured when the objects were present. The authors argued these gestures may have facilitated lexical memory for relevant words. However, it is important to note that imagery is used even when perceiving visual stimuli. When imagery co-occurs with perception of the object being represented, it plays a role in perceptual processing and identification (West, Morris, & Nichol, 1985, p. 14). It may be the case that rather than facilitating lexical memory, the gestures may have been supporting the imagery that was co-occuring with visual perception, though the two functions are not mutually exclusive. Thirdly, gesture rate was higher when describing noncodable objects in comparison to codable objects.

The authors argued that these objects placed a stronger load upon visuospatial working memory. Finally, speech rate was significantly lower when gesture was restricted in comparison to when gesture production was allowed. Since speech referred to visuospatial objects, gesture restriction may have weakened imagery generation in this task, leading to more dysfluencies in speech.

Smithson and Nicoladis (2014) tested whether or not gesture rate is systematically affected when the demands on imagistic resources are manipulated. Participants in this study watched two cartoon clips and then were asked to tell an experimenter what they remembered from these videos. For the retelling component of this study, participants were randomly assigned to wear video glasses with either a simple or complex moving image on the lenses (note that these images were unrelated to the cartoons). Complex visuospatial images are thought to be more disruptive to visuospatial processing than simple visuospatial images (Logie, 1986). The researchers hypothesized that if gesture plays a functional role in activating visuospatial imagery, then participants should gesture more while relaying a narrative while viewing a complex moving image than a simple moving image. The results revealed that participants in the complex visuospatial distractor condition produced significantly more gestures than participants in the simple visuospatial distractor condition. Smithson and Nicoladis (2014) concluded that these results lend support to the claim that gesture production facilitates imagery generation.

Gesture production may enhance recall for imagistic information. In a study investigating whether or not encouraging gesture production can enhance children's recall for experienced events, Stevanoni and Salmon (2005) asked 6 and 7 year old children to participate in an event titled 'a visit to the pirate'. Two weeks later their recall of the event was assessed in an interview. Children were assigned to one of four interview conditions: (1) a gesture instructed condition where they were instructed to both demonstrate and describe what happened during the event, (2) a gesture-modelled condition where they viewed the modeling of gesture production by the interviewer and were instructed to describe what happened during the event, (3) a gesture-allowed condition where children were only asked to describe what happened during the event, and (4) a gesture-restricted condition where children were prevented from gesturing when asked to describe what happened during the event. In order to prevent children from gesturing in the gesture-restricted condition, they were instructed to wear a special apron that was tied at the front. The children were asked to keep their hands in the special memory apron throughout the interview. The children in the gesture instructed condition gestured significantly more than children in the gesture-modelled and gesture-allowed conditions. Children in the gesture instructed condition verbally reported more correct information than children in the gesture-modelled, gesture-allowed, and gesture restricted conditions. When combining information conveyed in both gesture and speech, children in the gesture-instructed condition reported more than double the information reported by children in the gesture-restricted condition. The authors proposed three possible explanations for these results. Firstly, instructing gesture production may facilitate the re-enactment of the activities that took place during the event and this may facilitate retrieval. Secondly, gestures may serve as props for children and this may reduce processing demands thereby facilitating retrieval strategies. Finally, instructing gesture production may have enhanced task motivation. Participants in this condition may have been more strongly engaged in the task and therefore may have conveyed more information (Stevanoni & Salmon, 2005).

The influence of gesture production on strategy use. According to Broaders, Wagner-Cook, Mitchell, and Goldin-Meadow (2007) gestures can play an important role in revealing implicit knowledge. Broaders et al. (2007) were interested in determining whether encouraging gesture production could play a beneficial role in helping children to express implicit mathematical knowledge. Children were asked to solve two sets of six mathematical equivalence problems. Children were asked to solve the first six problems and explain how they solved them to an experimenter without any instructions regarding their hand movements. For the final six problems, participants were asked to solve the problems and to explain how they solved each of them while being assigned to one of the following between-subjects conditions: (1) no instructions regarding their hand movements (2) instructions to move their hands or (3) instructions to not move their hands. The number of new strategies expressed by children (i.e. the number of strategies that were not expressed by children while they solved the first six problems) was assessed. The results revealed that participants in the gesture encouraged condition added significantly more problem-solving strategies in comparison to children who were assigned to the control and gesture discouraged conditions. Importantly, the majority of these strategies were correct. This result led to a second study in which Broaders et al. (2007) explored whether children who were encouraged to gesture would be more receptive to instruction. In this study, a different group of participants completed six problems on paper at a desk and subsequently solved 6 problems and explained their reasoning to the problems while either being instructed to gesture or being discouraged from gesturing. Subsequently, children were given a lesson in mathematical equivalence. Children who were encouraged to gesture solved significantly more problems correctly in a posttest after the math instruction in comparison to children who were discouraged from gesturing.

Alibali, Spencer, Knox, and Kita (2011) investigated whether or not gesture production can influence the strategies that individuals use when completing problem-solving tasks. This was investigated in a study addressing gesture production during a task requiring participants to predict gear movements. There tends to be a transition of strategy use as individuals gain experience with this task. Initially, participants tend to rely on perceptual-motor strategies and with experience they tend to transition to the use of abstract strategies. Perceptual-motor strategies often involve the depiction of the movements of the gears, while abstract strategies require the participant to reason based on rules (e.g., determining whether the number of gears is odd or even to answer the question). The authors predicted that participants would be more likely to adopt an abstract strategy when gestures were restricted. Participants were randomly assigned to a gesture restriction or gesture allowed condition. In the gesture restriction condition, participants wore gloves that were attached to a board with Velcro. In the gesture allowed condition, participants' feet were strapped onto a board with Velcro straps. The foot straps were used in the gesture allowed condition in order to ensure that the level of strangeness was kept relatively constant across conditions. Participants were separated from the experimenter by an opaque screen. Participants solved six gear rotation problems while thinking aloud. The results indicated that participants in the gesture-restricted condition used abstract strategies more often than those in the gesture-allowed condition. Additionally, participants who were in the gesture-allowed condition who did not gesture tended to use abstract strategies more often than individuals who spontaneously gestured. In order to ensure that talking aloud through these problems was not responsible for influencing strategies, a second study was conducted in which participants solved the gear problems silently while either being allowed to gesture or restricted from gesturing. Participants in the gesture-allowed condition completed the problems with no restraints on their bodies, while the participants in the restricted condition were asked to sit on their hands. The percentage of correct solutions was not significantly different according to condition. Additionally, participants in the gesture-restricted condition were more likely to use abstract strategies than participants in the gesture-allowed condition. This study also replicated the finding that participants who were in the gesture-allowed condition who did not gesture, tended to use abstract strategies more often than individuals who spontaneously gestured. The authors note that in this study the use of gestures made it less likely that the participants would adopt the more efficient abstract strategy. They suggest that gestures play an important role in strategy choice by highlighting and structuring information in a perceptual-motor manner.

The Present Research

The studies reviewed suggest that gesture production may encourage the reliance upon a visuospatial code. The visuospatial code consists of information stored imagistically (Paivio, 1986). When gestures were allowed or encouraged in the studies reviewed, participants demonstrated higher levels of imagery in speech (Rimé et al., 1984), a faster speech rate when referring to visuospatial content (Morsella et al., 2004; Rauscher et al., 1996), and they remembered more information from an experienced event (Stevanoni & Salmon, 2005). More convincingly, when allowed to gesture people tended to adopt a perceptual-motor strategy instead of an abstract strategy in a problem-solving task (Alibali et al., 2011).

The studies reviewed also suggest that gesture restriction may encourage the reliance upon a verbal code. Verbal representations tend to be superior with respect to the representation of abstract knowledge (Kosslyn 1988; Kosslyn et al., 1995a) and tend to be stored in a sequential manner (Sadoski et al., 1991). When movement was restricted in the studies reviewed, participants tended to show lower levels of imagery in speech in comparison to when movement was allowed (Rimé et al., 1984). Furthermore, when speech referred to non-visuospatial content speakers tended to speak more quickly when they were restricted from gesturing (Rauscher et al., 1996). More persuasively, when restricted from gesturing, people tended to adopt an abstract strategy instead of a perceptual-motor strategy in a problem-solving task (Alibali et al., 2011).

The purpose of the present research was to determine whether gesture production has an effect upon whether individuals tend to rely upon a verbal or visuospatial code to complete a reading comprehension task and whether gesture production influences reading comprehension outcomes.

It is important to note that gesture production may not only influence reliance upon verbal and visuospatial codes in a reading comprehension task, but may also influence reliance upon intrinsic motivation. Research suggests that gesture production may be associated with motivation in other tasks. For example, in a study by Kelly, Byrne, and Holler (2011) the stakes of communication were manipulated and gesture production was assessed. When the stakes of communication were high, participants were found to gesture significantly more than when the stakes of communication were low. This research suggests that gesture rate may increase when participant task motivation is high. Whether gesture production influences the reliance upon intrinsic motivation in a reading comprehension task has yet to be investigated.

Both intrinsic motivation and cognitive factors tend to contribute uniquely to reading comprehension (Anmarkrud & Bråten, 2009; Taboada, Tonks, Wigfield, & Guthrie, 2009). The focus of the current study was the investigation of whether gesture production influences the reliance upon a verbal or visuospatial code, but it was also of interest to explore the possibility that gesture production may influence the reliance upon intrinsic motivation.

The present research will shed light upon the reading comprehension processes that occur among both adults and children. Study 1 will address reading comprehension among adults, while Study 2 will address reading comprehension among children.

CHAPTER II

Introduction to Study 1

Two Grimms' tales (Manheim, 1977) were used as the reading passages in this study: an easy reading passage called 'Old Sultan', and a standard reading passage called 'The Owl'. These passages were classified as 'easy' and 'standard' by previous researchers who conducted reading comprehension research with adults (Weaver & Bryant, 1995). Participants were assigned to a control condition, a gesture encouraged condition, a gesture restricted condition, or a written condition. Each reading passage was divided into three paragraphs. After each paragraph participants were asked to tell the experimenter (in the control, gesture encouraged, and gesture restricted conditions) or write down (in the written condition) what they remembered. When participants had finished summarizing each of the paragraphs they were asked to complete a fill-in-the-blank assessment of the reading passage. Following this, they were asked to provide a comprehensive summary of everything that they could remember from the entire reading passage.

According to Woolley (2011), the primary difference between a text-based and situation model is that the text-based model depends upon the generation of few inferences, whereas the situation model depends upon the generation of many inferences. The reading comprehension assessments used in this study were divided into those that required the generation of few or no inferences and those that required the generation of many inferences. These will be referred to as Low Inference and High Inference assessments, respectively. The Low Inference measures consisted of fill-in-the-blank questions that required the recall of verbatim information, and a rating of how many propositional units were recalled from the reading passage in the final comprehensive summary. The High Inference measures were ratings of a variety of dimensions of the final comprehensive summary by a trained coder. High Inference measures included ratings of the number of added details to the narrative upon recall (i.e., additional information which did not contradict the information presented within the reading passage) and a rating of the number of recall errors that were made upon recall (i.e., information which did contradict the information presented within the reading passage). High Inference ratings were not contingent upon the recall of verbatim information. Intrinsic motivation was assessed by the reading efficacy, reading challenge, reading curiosity, and reading involvement subscales of the Motivations for Reading Questionnaire (Wigfield & Guthrie, 1997), verbal and visuospatial working memory were assessed by the Automated Working Memory Assessment (Alloway, 2007), vocabulary was assessed by the Peabody Picture Vocabulary Task (Dunn & Dunn, 1997), and decoding speed was measured as well (these measures will be described in greater detail in the Methods section).

In line with the literature that has been reviewed, three sets of analyses were conducted in order to determine whether gesture production influences the adoption of a verbal or visuospatial code in a reading comprehension task: (1) predictors of reading comprehension were evaluated across experimental conditions, (2) performance on Low Inference and High Inference measures of reading comprehension were compared across experimental conditions, and (3) performance on Low Inference and High Inference measures of reading comprehension were compared across individuals who gestured spontaneously and who did not gesture spontaneously in the Control condition.

Part 1: Predictors of reading comprehension according to experimental condition

Hypothesis 1: *When gestures are encouraged and when they are free to be produced spontaneously, the visuospatial working memory measure will be the best predictor of reading comprehension*

The visuospatial code relies upon imagistic information (Paivio, 1986). Research concerning gesture production suggests that it may encourage the use of the visuospatial code. For example, gesture production tends to be associated with higher levels of imagery in speech (Rimé et al., 1984) and when people gesture they have been shown to adopt a perceptual-motor strategy over an abstract strategy in a problem-solving task (Alibali et al., 2011). It was predicted that among participants who were free to gesture or encouraged to gesture, reading comprehension outcomes would be most significantly determined by the visuospatial working memory measure.

Hypothesis 2: *When gestures are restricted, factors associated with a verbal strategy of encoding (i.e., vocabulary and verbal working memory measures) will be the best predictors of reading comprehension*

The verbal code stores information in a sequential manner (Sadoski et al., 1991) and is thought to represent abstract knowledge more effectively than the visuospatial code (Kosslyn 1988; Kosslyn et al., 1995a). Research suggests that being restricted from moving encourages the use of a verbal code. When movement is restricted, people tend to show lower levels of imagery in their speech (Rimé et al., 1984) and additionally, when restricted from moving, people have been shown to adopt an abstract strategy rather than a perceptual motor strategy in a problem-solving task (Alibali et al., 2011). If it is the case that the restriction of gesture production encourages the use of a verbal code, then reading comprehension outcomes when people are restricted from producing gestures should be largely determined by factors associated with the verbal code. It was predicted that in the gesture-restricted condition, reading comprehension outcomes would be most significantly determined by vocabulary and verbal working memory measures, rather than by the visuospatial working memory measure.

Part 2: Performance on Low Inference and High Inference measures of reading comprehension according to experimental condition

Hypothesis 3: *Individuals who are restricted from moving will perform better on Low Inference assessments*

The verbal text-based models are thought to retain verbatim information that is contained within the text and visuospatial situation models are not thought to retain verbatim information (Pearson & Johnson, 1978; Stull & Mayer, 2007). Evidence in support of this has been generated. For example, when assessing the effectiveness of verbal and visuospatial interventions, Johnson-Glenberg (2000) found that the verbal group tended to be better at recalling "factual, linear, highly verbal, text-explicit information" (p.780). If gesture restriction encourages the reliance upon a verbal code, then participants in this condition should excel on Low Inference reading comprehension measures. These measures include fill-in-the-blank assessments and also the proposition inclusion assessment.

Hypothesis 4: *Individuals who are encouraged to gesture will perform better on High Inference assessments*

The visuospatial situation models consist of elaborative inferences and link personal knowledge with the text information (Pearson & Johnson, 1978; Stull & Mayer, 2007). Evidence supporting this has been generated. When assessing the effectiveness of verbal and visuospatial interventions, Johnson-Glenberg (2000) found that the visuospatial group tended to excel at generating implicit open-ended questions and creating inferences. If gesture production encourages the reliance upon a visuospatial code, then participants who gesture should excel on High Inference reading comprehension assessments. These measures include ratings of elements of the narrative such as Summary integration and Added details.

Part 3: Performance on Low Inference and High Inference measures of reading comprehension according to whether individuals gestured spontaneously or not

Hypothesis 5: *Iconic gesture production is negatively associated with the measure of verbal working memory and positively associated with the measure of visuospatial working memory*

Researchers have suggested that individuals with strong spatial skills and low verbal skills tend to have high gesture rates (Hostetter & Alibali, 2007). It was hypothesized that individuals who gestured spontaneously would have higher visuospatial working memory scores and lower verbal working memory scores in comparison to individuals who did not gesture spontaneously.

Hypothesis 6: *Individuals who do not gesture spontaneously will perform better on Low Inference assessments than individuals who gesture spontaneously*

Recall that Johnson-Glenberg (2000) found that the verbal group tended to be better at recalling "factual, linear, highly verbal, text-explicit information" (p.780). If individuals who do not gesture spontaneously tend to rely upon a verbal code, it is predicted that they would outperform individuals who gesture spontaneously on Low Inference measures of reading comprehension.

Hypothesis 7: *Individuals who gesture spontaneously will perform better on High Inference assessments than individuals who do not gesture spontaneously*

Recall that Johnson-Glenberg (2000) found that the visuospatial group tended to excel at generating implicit open-ended questions. If individuals who gesture spontaneously tend to rely upon a visuospatial code, it is predicted that they would outperform individuals who do not gesture spontaneously on High Inference assessments.

Method

Participants

One hundred and twenty participants were recruited from the University of Alberta in Edmonton, Alberta. The participants were all English monolinguals and ranged in age from 16 to 23 years (41 males and 79 females). Participants were randomly assigned to one of four experimental conditions for the summary and narrative production component of the study: (1) *Gesture condition*: Participants in this condition were encouraged to use meaningful hand and arm movements as they relayed their narratives, (2) *Restricted condition*: Participants in this condition were restricted from moving their hands and arms as they relayed their narratives, (3) *Control condition*: Participants were not provided with any instructions regarding their movements in this condition, (4) *Written condition*: Participants in this condition wrote their narrative summaries using pen and paper. Outliers were identified as having values with $Z > \pm 3.29$ (99.9^{th} percentile) (Tabachnick & Fidell, 2007). Six participants were removed completely from the analyses. One participant did not complete the vocabulary measure due to time constraints. This participant was only excluded from analyses involving the vocabulary measure. One participant did not provide a final summary for the standard narrative passage due to experimenter error. This participant was only excluded from analyses involving ratings of this summary. Four participants in the Written condition did not complete the reading comprehension task using the standard narrative passage due to time restraints. These participants were excluded only from analyses involving the standard reading passage.

Materials

Reading passages. A practice reading and two test reading passages of increasing difficulty were selected for use with adults. All reading passages were from *Grimms' tales for young and old* (Manheim, 1977). A paragraph from ‘The Mouse, the Bird, and the Sausage’ was used as the practice reading passage. The test reading passages have been used in previous research concerning reading comprehension among adults (Weaver & Bryant, 1995). With respect to the test reading passages, the easy narrative reading was titled ‘Old Sultan’, and the standard narrative reading was titled ‘The Owl’. All of the narratives were presented on paper.

Each of the test reading passages had a speaking head icon after every main paragraph (a total of 3 per passage) in order to indicate where participants were to stop reading and provide a summary. For each reading passage, there were fill-in-the-blank reading comprehension questions to complete (1 for the practice passage and 10 for the test passages). See Appendix A for all reading comprehension questions used with adults.

Vocabulary assessment. In order to assess receptive vocabulary, the Peabody Picture Vocabulary Test – Third Edition (PPVT – IIIA) (Dunn & Dunn, 1997) was used.

Motivations for Reading Questionnaire (MRQ). Wigfield and Guthrie (1997) developed the Motivations for Reading Questionnaire (MRQ). This questionnaire assesses both extrinsic and intrinsic motivation to read. Research has shown that intrinsic motivation is positively associated with reading comprehension while extrinsic motivation is negatively associated with reading comprehension (Wang & Guthrie, 2004). Additionally, strong and weak readers differ with respect to their levels of intrinsic but not extrinsic reading motivation (Lau & Chan, 2003). Intrinsic motivation was the focus in the current research study. As a result, only 20 items were included from the MRQ representing four different facets of intrinsic motivation: reading efficacy, reading challenge, reading curiosity, and reading involvement. Psychometric properties for the MRQ are based on the 53-item questionnaire. The reliabilities range from .43 to .81 (Wigfield & Guthrie, 1997). Additionally, factor analyses shows construct validity supporting eleven factors (4 of which were used in the current study) (Wigfield & Guthrie, 1997). Other researchers have found support for the 11-factor model among late elementary school children with a confirmatory fit index of .90 suggesting a good model fit (Unrau & Schlackman, 2006).

Automated Working Memory Assessment (AWMA). The Automated Working Memory Assessment is a standardized tool that is used to assess verbal and visuospatial working memory resources (Alloway, 2007). This assessment tool is computerized. The scoring on the AWMA is automated and the sequence of all of the tasks is pre-set. Participants completed one visuospatial short-term memory, one visuospatial working memory and two verbal working memory tasks. Since the focus of this research was on working memory, performance on the visuospatial short-term memory task was not included in the analyses. Additionally, performance on one of the verbal working memory tasks ('Counting Recall') was not included in the analyses since this task presents information in a visuospatial format and therefore may have led some participants to adopt a visuospatial rather than verbal strategy. In the current study, the Odd-One-Out task was used as the measure of visuospatial working memory and the Listening Recall task was used as the measure of verbal working memory.

(1) *Visuospatial working memory task (Odd-One-Out):* In this task participants are shown 3 boxes with simple shapes presented simultaneously in each of them. Two of the shapes are identical while the third does not belong. Participants are asked to point to the shape that does not belong to the set when the shapes appear on the screen. When the shapes disappear from the screen they are asked to identify which box the odd-one-out shape appeared in. As the number of sets of shapes increases, the task becomes progressively more difficult since participants have to remember the location of each of the odd-one-out shapes in the same order as they were presented. The memory score used in the analyses reflects participant accuracy on the recall of odd-one-out shapes in the same order that they were presented.

(2) *Verbal working memory task (Listening Recall):* In this task, participants hear a sentence and they are asked to identify whether the sentence is true or false (based on the semantic content of the sentence). After that, participants are asked to remember the last word of the sentence that was presented. This task becomes progressively difficult as the number of sentences presented increases, since participants must remember a larger number of last words in the exact same order that they were heard. The memory score used in the analyses reflects participant accuracy for recalling the last words of the sentences in the same order that they were presented.

Gesture restriction device. In order to ensure that participants in the Gesture Restricted condition did not move their hands and arms while they were relaying their narratives, their hands were restricted. This restriction of movement was accomplished by asking participants to put on a pair of ski gloves. These ski gloves were then affixed to the armrests of a chair using Velcro fasteners. This arrangement restricted movements of the hands.

Procedure

Informed consent was obtained from all adults who participated in this study. Additionally, participants completed an information form which asked them to provide information concerning their age, language background, birth date, and gender before completing the study tasks. The participants completed the session as follows:

1) *Reading comprehension task:* Participants initially completed a practice reading passage. They were asked to stop reading when they came to a speaking head icon. At this point participants in the Gesture condition were asked: “Can you tell me everything that you can remember from that paragraph while trying to move your hands and arms in meaningful ways as you tell me?”; participants in the Restricted condition were asked “Can you tell me everything that you can remember from that paragraph while trying not to move?” (they relayed their summaries with

their hand movements restricted by ski gloves); participants in the Control condition were asked "Can you tell me everything that you can remember from that paragraph?"; and participants in the Written condition were asked "Can you write down everything that you can remember from that paragraph?". After reading to the end of the practice reading passage they were asked to complete one fill-in-the-blank question. Subsequently, they were asked to relay a comprehensive summary of the entire reading passage. Participants in the Gesture condition were asked: "Can you tell me everything that you can remember from that entire story while trying to move your hands and arms in meaningful ways as you tell me?"; participants in the Restricted condition were asked "Can you tell me everything that you can remember from that entire story while trying not to move?" (they relayed their summaries with their hand movements restricted by ski gloves); participants in the Control condition were asked "Can you tell me everything that you can remember from that entire story?"; and participants in the Written condition were asked "Can you write down everything that you can remember from that entire story?". The same procedure was followed for the easy and standard reading passages with the only exception being that 10 fill-in-the-blank reading comprehension questions were asked.

2) *AWMA*: Participants completed the memory assessments. The experimenter was responsible for identifying whether participant responses were correct or incorrect. In order to effectively do this, the experimenter placed the answer manual in a location that was visible only to her. The experimenter used the forward key on the computer keypad for correct responses and the backwards key on the computer keypad for incorrect responses. When participants received three errors in a testing block the program exited automatically, providing the experimenter with the relevant score. These were the only automated components of the program.

3) *PPVT:* Participants completed the vocabulary assessment. The experimenter said a word and the participant was shown four black and white pictures. The participant was asked to point to the picture that best corresponded to the word that the experimenter said.

4) *MRQ:* Participants answered 20 questions in Likert scale format regarding their intrinsic motivations to read.

Speech and gesture coding. A native English speaker transcribed all participant narratives in orthographic words. CLAN transcription software (MacWhinney, 2000) was used to determine the number of word types and tokens participants used to tell each narrative. Restarts and hesitations in speech were both counted in the number of word tokens. Gesture production was coded using the classification system that has been proposed by McNeill (1992; 2005). Only iconic gestures were coded since this type of gesture was the focus of the current study. Iconic gestures are hand and arm movements that convey semantic information that is strongly connected to the information expressed in speech (McNeill, 2005). An iconic gesture rate was calculated for each participant narrative by taking the number of iconic gestures divided by the total number of word tokens multiplied by 100. The number of iconic gestures was divided by the total number of word tokens in order to control for individual differences in story length. Additionally, the ratio of gestures to word tokens was multiplied by 100 for ease of interpretability.

Narrative evaluations. An independent coder used the rating scale in Appendix B in order to evaluate participant comprehensive narratives. The coder evaluated the narratives based on the transcriptions from the comprehensive narratives produced by participants. The coder evaluated the narratives on seven dimensions: (1) Main idea construction, (2) Summary integration, (3) Degree of understanding, (4) Information conveyed, (5) Listener enjoyment, (6) Added details, and (7) Recall errors. Questions 1 and 2 were adapted from previous work from

Kolic-Vehovec et al. (2011), question 3 was adapted from previous work by Thiede and Anderson (2003), and questions 4-7 were developed for the current study. Additionally, the coder evaluated the Accuracy for proposition inclusion. The standard reading passage was divided into 102 unique propositions. Participants were evaluated on whether they included each of these propositions (they received 1 point for the inclusion of each proposition in their final narrative summary). For all narrative evaluations, the coder only evaluated the standard reading passage.

Decoding speed. Decoding speed was the total time required for participants to read all three paragraphs in the standard narrative passage (measured in seconds).

Results

Before performing analyses to investigate the seven hypotheses, comparisons were conducted to assess whether the independent variables differed significantly across conditions and across spontaneous gesturers and non-gesturers. Additionally, comparisons were conducted to determine whether summary length differed significantly across conditions and across spontaneous gesturers and non-gestures since this may influence reading comprehension. The purpose of these comparisons was to identify any factors that needed to be statistically controlled when investigating the hypotheses.

Comparisons across experimental conditions

Independent measures

Descriptive statistics and one-way ANOVAs for age, memory, motivation, decoding, and vocabulary can be found in Table 1. The one-way ANOVAs indicate that none of these variables differed significantly according to experimental condition. As a result, none of these variables were statistically controlled in subsequent analyses. See Appendices D and E for information regarding the subtests of the MRQ and the individual paragraph timings for decoding speed, respectively.

Table 1

Average (SD) independent variable measures according to experimental condition

	Range	Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=29)	ANOVA
Age	16 - 23	18.39 (1.03)	18.93 (1.68)	18.30 (1.32)	18.62 (1.18)	F (3, 110) = 1.304 p = n.s., η_p^2 = 0.034
Gender		8 males 20 females	10 males 20 females	10 males 17 females	12 males 17 females	
Verbal WM	8 - 28	18.11 (3.77)	16.87 (4.44)	16.52 (3.95)	16.55 (4.12)	F (3, 110) = 0.937 p = n.s., η_p^2 = 0.025
Visuo WM	15 - 38	27.25 (5.41)	25.93 (3.86)	24.37 (5.15)	25.28 (4.80)	F (3, 110) = 1.748 p = n.s., η_p^2 = 0.046
Total MRQ	6.73 – 15.67	12.35 (1.97)	11.95 (1.82)	11.78 (2.04)	11.49 (1.95)	F (3, 110) = 0.969 p = n.s., η_p^2 = 0.026

		Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=25)	
Standard decoding speed	186 - 351	247.89 (30.12)	252.17 (33.59)	255.56 (42.87)	248.64 (42.01)	*F* (3, 106) = 0.242 p = n.s., η_p^2 = 0.007
		Control (N=28)	Gesture (N=30)	Restricted (N=26)	Written (N=29)	
PPVT	155 - 193	182.18 (6.75)	180.43 (7.06)	179.50 (9.97)	178.45 (9.05)	*F* (3, 109) = 1.036 p = n.s., η_p^2 = 0.028

Summary tokens and types

Intermittent summary types did not differ across groups. Intermittent summary tokens did differ significantly. Pairwise Bonferroni-correct post-hoc t-tests indicated that participants in the Written condition used significantly fewer intermittent summary tokens than participants in the Gesture condition ($M_D = 233.05, SE_D = 47.25, p < .001$), the Control condition ($M_D = 153.77, SE_D = 48.43, p < .05$), and the Restricted condition ($M_D = 168.44, SE_D = 48.43, p < .01$). All other comparisons were non-significant (see Table 2) (Note: the D subscript is used since these values refer to difference scores).

Final summary types differed significantly across groups. Pairwise Bonferroni-correct post-hoc t-tests indicated that participants in the Written condition used fewer final summary types than participants in the Gesture condition ($M_D = 46.16, SE_D = 16.26, p < .05$). All other comparisons were non-significant. Final summary tokens differed significantly across groups. Pairwise Bonferroni-correct post-hoc t-tests revealed that participants in the Written condition used fewer final summary tokens than participants in the Gesture condition ($M_D = 227.33, SE_D = 48.77, p < .001$), the Control condition ($M_D = 176.86, SE_D = 49.99, p < .01$), and the Restricted condition ($M_D = 166.98, SE_D = 49.99, p < .01$). All other comparisons were non-significant (see Table 2).

Table 2

Average (SD) word types and tokens used throughout intermittent and final summaries across experimental conditions

	Range	Control (C) (N=27)	Gesture (G) (N=30)	Restricted (R) (N=27)	Written (W) (N=25)	ANOVA
Intermittent summary types	65 - 324	172.74 (56.64)	193.20 (62.26)	177.70 (57.12)	156.56 (38.63)	F (3, 105) = 2.067 p = n.s., $\eta_p^2 = 0.056$
Intermittent summary tokens	132 - 1043	432.85 (188.10)	512.13 (211.57)	447.52 (173.60)	279.08 (90.24)	F (3, 105) = 8.484 $p < .001$, $\eta_p^2 = 0.195$ (W < G, C, R)
Final summary types	43 - 301	150.93 (63.37)	167.80 (65.29)	153.11 (64.39)	121.64 (42.09)	F (3, 105) = 2.762 $p < .05$, $\eta_p^2 = 0.073$ (W < G)
Final summary tokens	64 - 927	384.70 (198.61)	435.17 (210.00)	374.81 (188.47)	207.84 (85.95)	F (3, 105) = 7.900 $p < .001$, $\eta_p^2 = 0.184$ (W < G, C, R)

Word tokens and types throughout intermittent and final summaries were all strongly correlated (see Table 3).

Table 3

Pearson correlations between intermittent (I) and final (F) summary tokens and types (assessed among all participants) (df = 107)

Measure	2	3	4
1) Types (I)	0.931**	0.929**	0.891**
2) Tokens (I)	-	0.888**	0.943**
3) Types (F)	-	-	0.955**
4) Tokens (F)	-	-	-

**p < .001

In the analyses conducted to assess hypotheses 3 and 4 (which tested Low and High Inference reading comprehension outcomes, respectively, across conditions), intermittent summary tokens were statistically controlled. The intermittent summary rather than final summary tokens were chosen to be statistically controlled since the intermittent summary length could influence performance upon the fill-in-the-blank measures of reading comprehension. Final summary length could not since the final summaries were relayed after the completion of the fill-in-the-blank questions.

Comparisons across spontaneous gesturers and non-gesturers in the Control condition

Independent measures

Within the Control condition, 2/3 of the participants spontaneously gestured at least once. Descriptive statistics and independent samples t-tests for age, working memory, motivation, decoding speed, and vocabulary can be found in Table 4. The t-tests indicate that only the motivation scores differed according to whether or not participants produced iconic gestures spontaneously. More specifically, individuals who gestured spontaneously had higher motivation scores on the MRQ than individuals who did not gesture spontaneously. Motivation was therefore statistically controlled in the analyses used to assess hypotheses 6 and 7 (which tested Low and High Inference reading comprehension outcomes, respectively, across spontaneous gesturers and non-gesturers).

Table 4

Average (SD) age, memory, motivation, decoding, and vocabulary measures according to spontaneous gesturers and non-gesturers in the Control condition

Measure	Non-Gesturers (N=9)	Gesturers (N=18)	Independent samples t-test
Age	18.67 (1.50)	18.22 (0.73)	*t* (9.953) = 0.840, *p* = n.s., *Cohen's d* = 0.381
Gender	3 males 6 females	4 males 14 females	
Verbal WM	18.56 (3.05)	18.11 (4.14)	*t* (25) = 0.285, *p* = n.s., *Cohen's d* = 0.124
Visuo WM	25.56 (4.25)	28.39 (5.78)	*t* (25) = -1.300, *p* = n.s., *Cohen's d* = -0.558
Total MRQ	11.03 (1.77)	13.06 (1.79)	*t* (25) = -2.784, *p* < .05, *Cohen's d* = -1.140
Standard decoding speed	244.89 (32.59)	249.83 (30.47)	*t* (25) = -0.389, *p* = n.s., *Cohen's d* = -0.157
PPVT	181.44 (7.58)	182.22 (6.56)	*t* (25) = -0.276, *p* = n.s., *Cohen's d* = -0.110

Summary tokens and types

Independent samples t-tests revealed that participants who gestured spontaneously used significantly more types and tokens in their intermittent and final summaries than participants who did not gesture spontaneously (see Table 5). As a result, intermittent word tokens were statistically controlled in the analyses used to assess hypotheses 6 and 7.

Table 5

Group comparisons of average (SD) word types and tokens throughout intermittent and final summaries

	Non-Gesturers (N=9)	Gesturers (N=18)	*t*-test
Intermittent summary types	133.56 (49.28)	192.33 (50.44)	t (25) = -2.875, $p < .01$, *Cohen's d* = -1.179
Intermittent summary tokens	308.44 (139.19)	495.06 (180.85)	t (25) = -2.711, $p < .05$, *Cohen's d* = -1.156
Final summary types	109.11 (55.20)	171.83 (57.61)	t (25) = -2.703, $p < .05$, *Cohen's d* = -1.112
Final summary tokens	253.56 (150.70)	450.28 (189.70)	t (25) = -2.705, $p < .05$, *Cohen's d* = -1.148

Summary of Analyses

The analyses were divided into three sections to investigate each of the seven hypotheses.

Part 1: Predictors of reading comprehension according to experimental condition

In part 1, forward multiple linear regression analyses were conducted to determine whether visuospatial working memory would be the best predictor of reading comprehension among participants in the Control and Gesture conditions (Hypothesis 1) and whether factors associated with a verbal strategy of encoding would be the best predictors of reading comprehension among participants in the Restricted condition (Hypothesis 2).

Part 2: Performance on Low Inference and High Inference measures of reading comprehension according to experimental condition

In part 2, multiple regression was performed to determine whether performance on Low Inference assessments (Hypothesis 3) and High Inference assessments (Hypothesis 4) differed significantly across groups and if so, to determine whether the inclusion of intermittent summary length would influence model fit. Since the experimental condition is a categorical variable and since the sample sizes were not equal across groups, dummy coding was used for these analyses. Each experimental group was assigned dummy values (*D1*, *D2, D3*): Control condition (0, 0, 0); Gesture condition (1, 0, 0); Restricted condition (0, 1, 0); and Written condition (0, 0, 1). The Control group was designated as the reference group.

Part 3: Performance on Low Inference and High Inference measures of reading comprehension according to whether individuals gestured spontaneously or not

The analyses conducted in Part 3 involved only participants in the Control group. In this section of the results, a multiple regression analysis was conducted to determine whether individual differences in working memory profiles predict gesture use (Hypothesis 5) and if so, to determine whether the inclusion of intermittent summary length and motivation would influence model fit (since these two variables differed significantly across spontaneous gesturers and non-gesturers).

Multiple regression was performed to determine whether performance on Low Inference assessments (Hypothesis 6) and High Inference assessments (Hypothesis 7) differed significantly across the spontaneous gesturers and non-gesturers and if so, to determine whether the inclusion of intermittent summary length and motivation would improve model fit. Dummy coding was used to code participants as spontaneous gesturers (1) or non-gesturers (0).

Note: See Appendix F for an alternative approach to the analyses.

Part 1: Predictors of reading comprehension according to experimental condition

Though specific hypotheses were generated regarding the predictors of reading comprehension outcomes in the different conditions, it was of interest to not only include the verbal (PPVT and verbal working memory) and visuospatial (visuospatial working memory) factors, but also to include the motivation factors. Gesture production may not only influence the reliance upon a verbal or visuospatial strategy, but may also be related to motivation (Kelly et al., 2011). It was of interest to explore this possibility. In order to investigate hypotheses 1 and 2, PPVT, motivation, verbal working memory, and visuospatial working memory measures were assessed as predictors of reading comprehension. See Appendix C for correlations between predictor and outcome variables according to experimental condition.

Hypothesis 1: *When gestures are encouraged and when they are free to be produced spontaneously, the visuospatial working memory measure will be the best predictor of reading comprehension*

In order to investigate whether visuospatial working memory would be the best predictor of reading comprehension outcomes among participants in the Control and Gesture encouraged conditions, forward multiple linear regression analyses were conducted.

Control condition

Low Inference measures

With respect to the Low Inference measures of reading comprehension, no significant predictors emerged for the Fill-in-the-blank (easy) assessment, however significant predictors emerged for the Fill-in-the-blank (standard) assessment and for the Accuracy for proposition inclusion score (see Table 6).

The curiosity subscale of the MRQ was a positive predictor of scores on the Fill-in-the-blank (standard) assessment. The curiosity subscale of the MRQ and the visuospatial working memory measure were positive predictors of Accuracy for proposition inclusion.

Table 6

Information concerning regression analyses for Low Inference measures among participants in the Control condition

Measure	*Adj.* R^2	*F test*	*Regression equation*	*Predictor(s)*
Fill-in-the-blank (standard)	0.315	$F(1, 26) = 13.421$ $p < .01$	$Y' = -0.611 + 2.202$(MRQ_Curiosity)	***MRQ_Curiosity:*** $\beta = .583$ $t(26) = 3.663$, $p < .01$
Accuracy for proposition inclusion	0.515	$F(2, 24) = 14.797$ $p < .001$	$Y' = 16.583 + 9.049$(MRQ_Curiosity) + 0.781(Visuospatial WM)	***MRQ_Curiosity:*** $\beta = .523$ $t(24) = 3.822$, $p < .01$ ***Visuospatial WM:*** $\beta = .498$ $t(24) = 3.641$, $p < .01$

High Inference measures

Significant predictors emerged for all High Inference measures (see Table 7). The visuospatial working memory measure and the curiosity subscale of the MRQ emerged as the most influential predictors. The visuospatial working memory measure was a positive predictor of Main idea construction, Summary integration, Understanding, Informative, and Enjoyment ratings; and a negative predictor of the Recall error rating. The curiosity subscale of the MRQ was a positive predictor of Main idea construction, Summary integration, Understanding, and Informative ratings; and a negative predictor of the Recall error rating. These results indicate that having strong visuospatial working memory abilities and high levels of intrinsic motivation for curiosity was advantageous for High Inference reading comprehension outcomes.

The challenge and efficacy subscales of the MRQ were significant predictors of some of the High Inference measures and had an adverse effect upon performance on the High Inference measures. The challenge subscale was a negative predictor of Informative ratings and a positive predictor of the Recall error rating. The efficacy subscale was a negative predictor of Summary integration.

The verbal working memory measure was a negative predictor of the Added details rating. However, it should be noted that when considering only those who gestured in the Control condition, the verbal working memory measure was no longer a predictor but rather motivation (involvement) was a significant and negative predictor of the Added details rating [Adjusted $R^2 = 0.347$, $F(1, 16) = 10.050$, $p < .01$].

Table 7

Information concerning regression analyses for High Inference measures among participants in the Control condition

Measure	*Adj.* R^2	*F test*	*Regression equation*	*Predictor(s)*
Main idea construction	0.240	F (2, 24) = 5.105, p < 0.05	Y' = -1.509 + 0.660(MRQ_Curiosity) + 0.059(Visuospatial WM)	***MRQ_Curiosity:*** β = .377 t (24) = 2.199, p < .05 ***Visuospatial WM:*** β = .374 t (24) = 2.186, p < .05
Summary integration	0.402	F (3, 23) = 6.836, p < .01	Y' = -0.403 + 0.936(MRQ_Curiosity) + 0.047(Visuospatial WM) – 0.443(MRQ_Efficacy)	***MRQ_Curiosity:*** β = .720 t (23) = 3.621, p < .01 ***Visuospatial WM:*** β = .397 t (23) = 2.616, p < .05 ***MRQ_Efficacy:*** β = -.466 t (23) = -2.341, p < .05
Understanding	0.479	F (2, 24) = 12.940, p < .001	Y' = -4.057 + 1.771(MRQ_Curiosity) + 0.115(Visuospatial WM)	***MRQ_Curiosity:*** β = .571 t (24) = 4.024, p < .001 ***Visuospatial WM:*** β = .407 t (24) = 2.873, p < .01
Informative	0.426	F (3, 23) = 7.422, p < .01	Y' = -5.430 + 3.911(MRQ_Curiosity) + 0.165(Visuospatial WM) –	***MRQ_Curiosity:*** β = .829 t (23) = 3.823, p < .01 ***Visuospatial WM:*** β = .387 t (23) = 2.591, p < .05

			1.883(MRQ_Challenge)	***MRQ_Challenge:*** $\beta = -.480$ $t\,(23) = -2.212, p < .05$
Enjoyment	0.122	$F\,(1, 25) =$ 4.614, $p < .05$	$Y' = 2.044 +$ 0.105(Visuospatial WM)	***Visuospatial WM:*** $\beta = .395$ $t\,(25) = 2.148, p < .05$
Added details	0.242	$F\,(1, 25) =$ 9.316, $p < .01$	$Y' = 4.086 - 0.141$(Verbal WM)	***Verbal WM:*** $\beta = -.521$ $t\,(25) = -3.052, p < .01$
Recall errors	0.619	$F\,(3, 23) =$ 15.088, $p < .001$	$Y' = 19.090 –$ 4.866(MRQ_Curiosity) – 0.215(Visuospatial WM) + 2.178(MRQ_Challenge)	***MRQ_Curiosity:*** $\beta = -.938$ $t\,(23) = -5.314, p < .001$ ***Visuospatial WM:*** $\beta = -.457$ $t\,(23) = -3.762, p < .01$ ***MRQ_Challenge:*** $\beta = .506$ $t\,(23) = 2.858, p < .01$

Gesture

Low Inference measures

Significant predictors emerged for all Low Inference measures in the Gesture condition (see Table 8). The involvement and efficacy subscales of the MRQ were the only significant predictors of Low Inference reading comprehension outcomes. More specifically, the involvement subscale of the MRQ was a positive predictor of the Fill-in-the-blank (easy) and Fill-in-the-blank (standard) scores, while the efficacy subscale of the MRQ was a positive predictor of the Fill-in-the-blank (standard) and Accuracy for proposition inclusion scores.

Table 8

Information concerning regression analyses for Low Inference measures among participants in the Gesture condition

Measure	*Adj. R^2*	*F test*	*Regression equation*	*Predictor(s)*
Fill-in-the-blank (easy)	0.296	F (1, 28) = 13.166, $p < .01$	Y' = 4.380 + 1.397(MRQ_Involvement)	***MRQ_Involvement:*** β = .566 t (28) = 3.629, $p < .01$
Fill-in-the-blank (standard)	0.373	F (2, 27) = 9.616, $p < .01$	Y' = 0.181 + 1.385(MRQ_Efficacy) + 0.784(MRQ_Involvement)	***MRQ_Efficacy:*** β = .540 t (27) = 3.657, $p < .01$ ***MRQ_Involvement:*** β = .313 t (27) = 2.123, $p < .05$
Accuracy for proposition inclusion	0.197	F (1, 28) = 8.115, $p < .01$	Y' = 49.386 + 6.626(MRQ_Efficacy)	***MRQ_Efficacy:*** β = .474 t (28) = 2.849, $p < .01$

High Inference measures

Significant predictors emerged for all High Inference measures in the Gesture condition with the exception of ratings of Summary integration, Enjoyment, and Added details (see Table 9). The efficacy subscale of the MRQ was the most influential predictor with respect to High Inference measures. More specifically, the efficacy subscale of the MRQ was a positive predictor of Main idea construction and Understanding; and was a negative predictor of the Recall error rating. The challenge subscale of the MRQ was a positive predictor of Informative ratings.

Table 9

Information concerning regression analyses for High Inference measures among participants in the Gesture condition

Measure	*Adj.* R^2	*F test*	*Regression equation*	*Predictor(s)*
Main idea construction	0.170	F (1, 28) = 6.948, $p < .05$	Y' = 0.962 + 0.482(MRQ_Efficacy)	***MRQ_Efficacy:*** β = .446 t (28) = 2.636, $p < .05$
Understanding	0.272	F (1, 28) = 11.810, $p < .01$	Y' = 0.903 + 1.276(MRQ_Efficacy)	***MRQ_Efficacy:*** β = .545 t (28) = 3.437, $p < .01$
Informative	0.143	F (1, 28) = 5.857, $p < .05$	Y' = 1.529 + 1.430(MRQ_Challenge)	***MRQ_Challenge:*** β = .416 t (28) = 2.420, $p < .05$
Recall errors	0.229	F (1, 28) = 9.615, $p < .01$	Y' = 9.798 – 1.933(MRQ_Efficacy)	***MRQ_Efficacy:*** β = -.506 t (28) = -3.101, $p < .01$

Summary of results relevant to hypothesis 1: It was hypothesized that among participants in the Control and Gesture conditions, the visuospatial working memory measure would be the best predictor of reading comprehension outcomes. Visuospatial working memory and motivation measures were the strongest predictors of Low Inference and High Inference reading comprehension outcomes in the Control condition. Motivation measures were the strongest predictors of Low and High Inference reading comprehension outcomes in the Gesture condition. These results lend only weak support to the hypothesis.

Hypothesis 2: *When gestures are restricted, factors associated with a verbal strategy of encoding (i.e., vocabulary and verbal working memory measures) will be the best predictors of reading comprehension*

Analyses were conducted in order to determine whether factors associated with a verbal strategy of encoding would be the best predictors of reading comprehension among participants in the Restricted condition. Forward multiple linear regression analyses were conducted to assess whether PPVT, motivation, verbal working memory, or visuospatial working memory measures were significant predictors of reading comprehension.

Restricted

Low Inference measures

Significant predictors emerged for the Fill-in-the-blank (standard) and Accuracy for proposition inclusion measures, but not the Fill-in-the-blank (easy) measure (see Table 10). PPVT was a positive predictor of both Fill-in-the-blank (standard) and Accuracy for proposition inclusion scores.

Table 10

Information concerning regression analyses for Low Inference measures among participants in the Restricted condition

Measure	*Adj.* R^2	*F test*	*Regression equation*	*Predictor(s)*
Fill-in-the-blank (standard)	0.204	F (1, 24) = 7.400, $p < .05$	$Y' = -3.749 + 0.061(\text{PPVT})$	***PPVT***: $\beta = .485$ t (24) = 2.720, $p < .05$
Accuracy for proposition inclusion	0.116	F (1, 24) = 4.268, $p = .050$	$Y' = 11.529 + 0.316(\text{PPVT})$	***PPVT:*** $\beta = .389$ t (24) = 2.066, $p = .050$

High Inference measures

No significant predictors emerged for Main idea construction, Summary integration, Understanding, Informative, Enjoyment, and Added detail ratings. However PPVT emerged a significant and negative predictor of the Recall error rating (see Table 11).

Table 11

Information concerning the regression analysis for the Recall error rating among participants in the Restricted condition

Measure	*Adj. R^2*	*F test*	*Regression equation*	*Predictor(s)*
Recall errors	0.158	F (1, 24) = 5.677, $p < .05$	$Y' = 22.420 - 0.102(\text{PPVT})$	***PPVT:*** $\beta = -.437$ t (24) = -2.383, $p < .05$

Summary of results relevant to hypothesis 2: It was hypothesized that among participants in Restricted condition, factors associated with a verbal strategy of encoding would be the best predictors of reading comprehension. Vocabulary was the only factor to emerge as a significant predictor of Low Inference and High Inference reading comprehension outcomes, lending support to this hypothesis.

Part 2: Performance on Low Inference and High Inference measures of reading comprehension according to experimental condition

Recall that each experimental group was assigned dummy values (*D1*, *D2, D3*): Control condition (0, 0, 0); Gesture condition (1, 0, 0); Restricted condition (0, 1, 0); and Written condition (0, 0, 1). In the current analyses, D_1 refers to the Gesture condition, D_2 refers to the Restricted condition, and D_3 refers to the Written condition.

Hypothesis 3: *Individuals who are restricted from moving will perform better on Low Inference assessments*

Low Inference reading comprehension outcomes across experimental conditions

Multiple regression was performed to investigate whether performance on the Low Inference assessments differed across experimental conditions and to determine whether the addition of intermittent summary length would influence model fit. For descriptive statistics across experimental conditions on the Low Inference assessments, see Table 12.

Table 12

Average (SD) Low Inference reading comprehension outcomes across experimental conditions

	Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=25)
Fill-in-the-blank (standard)	6.57 (1.89)	6.43 (1.57)	7.22 (1.34)	5.80 (1.47)
	Control (N=27)	Gesture (N=30)	Restricted (N=27)	Written (N=25)
Accuracy for proposition inclusion (standard)	67.28 (8.47)	68.23 (8.54)	69.09 (8.99)	66.56 (6.94)

Experimental condition influenced performance on the Fill-in-the-blank (standard) assessment but not Accuracy for proposition inclusion (standard). Since no dummy variables were entered into the regression equation predicting Accuracy for proposition inclusion, this rating will not be discussed.

Fill-in-the-blank (standard):

In the first step, the dummy variables were entered into the multiple regression using forward entry: D_1, D_2, and D_3. Only D_2 (Restricted condition) was included. This model was statistically significant F (1, 107) = 6.883, $p < 0.05$ and explained 5.2% of the variance in Fill-in-the-blank (standard) scores (see Table 13). After the entry of intermittent summary tokens, the model explained 15.4% of the variance in Fill-in-the-blank (standard) scores [F (2, 106) = 10.850, $p < .001$]. The addition of intermittent summary tokens explained an additional 10.9% of the variance in Fill-in-the-blank (standard) scores after controlling for the influence of D_2 [ΔR^2 = .109; F (1, 106) = 13.983, p < .001]. In the final model, both D_2 (β = .221, p < 0.05) and intermittent summary tokens (β = .332, p < .001) were statistically significant. In the final model the regression coefficient associated with D_2 was positive indicating that participants in the Restricted group performed significantly better on the Fill-in-the-blank (standard) assessment than participants in the Control group.

Table 13

Regression model predicting Fill-in-the-blank (standard) performance

	R	*R²*	*Adj. R²*	*ΔR²*	*B*	*SE*	*β*	*t*
Step 1	.246	.060	.052					
D_2					0.945	0.360	.246	2.624*
Step 2	.412	.170	.154	.109***				
D_2					0.851	0.341	.221	2.495*
Tokens (I)					0.003	0.001	.332	3.739***

*p < .05; **p < .01; ***p < .001

Summary of results relevant to hypothesis 3: It was hypothesized that individuals who were restricted from moving would demonstrate superior performance on Low Inference measures. This hypothesis was partially supported. Individuals in the Restricted condition performed significantly better than participants in the Control condition on the Fill-in-the-blank (standard) assessment, but performance on the Accuracy for proposition inclusion did not differ significantly across experimental conditions. When controlling for intermittent summary tokens, the Restricted condition advantage on the Fill-in-the-blank (standard) assessment persisted.

Hypothesis 4: *Individuals who are encouraged to gesture will perform better on High Inference assessments*

Multiple regression was performed to investigate whether performance on the High Inference assessments differed across experimental conditions and to determine whether the addition of intermittent summary length would influence model fit. For descriptive statistics across experimental groups on High Inference assessments, see Table 14.

Table 14

Average (SD) High Inference reading comprehension outcomes across experimental conditions

	Control (N=27)	Gesture (N=30)	Restricted (N=27)	Written (N=25)
Main idea construction	2.26 (0.86)	2.33 (0.66)	2.33 (0.73)	2.64 (0.57)
Summary integration	2.59 (0.64)	2.60 (0.56)	2.67 (0.48)	2.92 (0.28)
Understanding	4.81 (1.52)	4.53 (1.43)	4.85 (1.51)	5.20 (1.12)
Informative	6.04 (2.31)	5.80 (1.92)	6.19 (1.90)	6.96 (1.54)
Enjoyment	4.93 (1.44)	5.10 (1.83)	5.52 (1.19)	5.44 (1.58)
Added details	1.52 (1.01)	2.10 (1.27)	1.22 (1.12)	1.40 (1.15)
Recall errors	4.07 (2.54)	4.30 (2.34)	4.04 (2.41)	3.04 (1.86)

Experimental condition influenced all High Inference reading comprehension outcomes with the exception of ratings of Understanding and Enjoyment. The specific influences of experimental condition are explored in the following analyses.

Main idea construction:

The dummy variables were entered into the first step using forward entry. Only D_3 (Written condition) was included. This model was statistically significant F (1, 107) = 4.192, $p <$ .05 and explained 2.9% of the variance in Main idea construction ratings (see Table 15). After the entry of intermittent summary tokens, the model explained 42.3% of the variance in Main idea construction ratings. [F (2, 106) = 40.520, $p < .001$]. The addition of intermittent summary tokens explained an additional 39.6% of the variance in Main idea construction ratings after controlling for the influence of D_3 [$\Delta R^2 = .396$; F (1, 106) = 73.990, $p < .001$]. In the final model, both D_3 (β = .478, p < .001) and intermittent summary tokens (β = .690, p < .001) were statistically significant. In the final model, the regression coefficient associated with D_3 was positive indicating that participants in the Written condition had significantly higher Main idea construction ratings than participants in the Control condition.

Table 15

Regression model predicting Main idea construction rating

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.194	.038	.029					
D_3					0.319	0.156	.194	2.047*
Step 2	.658	.433	.423	.396				
D_3					0.786	0.132	.478	5.959***
Tokens (I)					0.003	0.000	.690	8.602***

*$p < .05$; **$p < .01$; ***$p < .001$

Summary integration:

In the first step, the dummy variables were entered using forward entry. Only D_3 (Written condition) was included. This model was statistically significant $F\ (1, 107) = 6.753$, $p < .05$ and explained 5.1% of the variance in Summary integration ratings (see Table 16). After the entry of intermittent summary tokens, the model explained 22.8% of the variance in Summary integration ratings. [$F\ (2, 106) = 16.980$, $p < .001$]. The addition of intermittent summary tokens explained an additional 18.3% of the variance in Summary integration ratings after controlling for the influence of D_3 [$\Delta R^2 = .183$; $F\ (1, 106) = 25.652$, $p < .001$]. In the final model, both D_3 ($\beta = .437$, $p < 0.001$) and intermittent summary tokens ($\beta = .470$, $p < .001$) were statistically significant. In the final model the regression coefficient associated with D_3 was positive, indicating that participants in the Written condition had significantly higher Summary integration ratings than participants in the Control condition.

Table 16

Regression model predicting Summary integration rating

	R	*R*^{2}	*Adj. R*^{2}	ΔR^2	*B*	*SE*	β	*t*
Step 1	.244	.059	.051					
D_3					0.291	0.112	.244	2.599*
Step 2	.493	.243	.228	.183				
D_3					0.521	0.111	.437	4.711***
Tokens (I)					0.001	0.000	.470	5.065***

*$p < .05$; **$p < .01$; ***$p < .001$

Informative:

In Step 1, the dummy variables were entered using forward entry. Only D_3 (Written condition) was included. This model was statistically significant $F\ (1, 107) = 4.762$, $p < .05$ and explained 3.4% of the variance in Informative ratings (see Table 17). After the entry of intermittent summary tokens, the model explained 45.1% of the variance in Informative ratings [$F\ (2, 106) = 45.426$, $p < .001$]. The addition of intermittent summary tokens explained an additional 41.9% of the variance in Informative ratings after controlling for the influence of D_3 [$\Delta R^2 = .419$; $F\ (1, 106) = 82.465$, $p < .001$]. In the final model, both D_3 ($\beta = .499$, $p < .001$) and intermittent summary tokens ($\beta = .710$, $p < .001$) were statistically significant. In the final model the regression coefficient associated with D_3 was positive, indicating that participants in the Written condition had significantly higher Informative ratings than participants in the Control condition.

Table 17

Regression model predicting Informative rating

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.206	.043	.034					
D_3					0.927	0.425	.206	2.182*
Step 2	.679	.462	.451	.419				
D_3					2.239	0.351	.499	6.376***
Tokens (I)					0.007	0.001	.710	9.081***

*$p < .05$; **$p < .01$; ***$p < .001$

Added details:

In Step 1, the dummy variables were entered using forward entry. Only D_1 (Gesture condition) was included. This model was statistically significant F (1, 107) = 8.664, $p < .01$ and explained 6.6% of the variance in Added details (see Table 18). After the entry of intermittent summary tokens, the model explained 6.7% of the variance in Added details [F (2, 106) = 4.857, $p < .05$]. The addition of intermittent summary tokens explained an additional 0.9% of the variance in Added details after controlling for the influence of D_1. This change was not statistically significant [$\Delta R^2 = .009$; F (1, 106) = 1.046, p = n.s.]. In the final model only D_1 (β = .245, $p < .05$) was statistically significant, indicating that participants in the Gesture condition had significantly higher added detail ratings than participants in the Control condition.

Table 18

Regression model predicting Added details rating

	R	R^2	Adj. R^2	ΔR^2	B	SE	β	t
Step 1	.274	.075	.066					
D_1					0.731	0.248	.274	2.943**
Step 2	.290	.084	.067	.009				
D_1					0.654	0.259	.245	2.525*
Tokens (I)					0.001	0.001	.099	1.023

*$p < .05$; **$p < .01$; ***$p < .001$

Recall errors:

In Step 1, the dummy variables were entered using forward entry. Only D_3 (Written condition) was included. This model was statistically significant F (1, 107) = 4.467, $p < .05$ and explained 3.1% of the variance in the Recall error rating (see Table 19). After the entry of intermittent summary tokens, the model explained 39.0% of the variance in the Recall error rating [F (2, 106) = 35.539, $p < .001$]. The addition of intermittent summary tokens explained an additional 36.1% of the variance in the Recall error rating after controlling for the influence of D_3 [$\Delta R^2 = .361$; F (1, 106) = 63.981, $p < .001$]. In the final model, both D_3 ($\beta = -.472$, $p < .001$) and intermittent summary tokens ($\beta = -.659$, $p < .001$) were statistically significant. In the final model the regression coefficient associated with D_3 was negative, indicating that participants in the Written group had lower Recall error ratings than participants in the Control group.

Table 19

Regression model predicting Recall error rating

	R	R^2	*Adj.* R^2	ΔR^2	B	SE	β	t
Step 1	.200	.040	.031					
D_3					-1.065	0.504	-.200	-2.114*
Step 2	.634	.401	.390	.361				
D_3					-2.508	0.439	-.472	-5.719***
Tokens (I)					-0.008	0.001	-.659	-7.999***

*$p < .05$; **$p < .01$; ***$p < .001$

Summary of results relevant to hypothesis 4: It was hypothesized that individuals who were encouraged to gesture would demonstrate advantages on the High Inference measures of reading comprehension. This hypothesis was not well supported by the results. Participants in the Gesture condition had significantly higher Added details ratings than participants in the Control condition; however, participants in the Written condition demonstrated the strongest advantages on the High Inference measures.

Part 3: Performance on Low Inference and High Inference measures of reading comprehension according to whether individuals gestured spontaneously or not

Hypothesis 5: *Iconic gesture production is negatively associated with the measure of verbal working memory and positively associated with the measure of visuospatial working memory*

It was predicted that the propensity to use gestures during a narrative task is associated with higher scores on the visuospatial working memory measure and lower scores on the verbal working memory measure. Since motivation scores and intermittent summary length were significantly different across gesturers and non-gesturers these factors were included in the linear regression model.

Verbal and visuospatial working memory measures were used as predictors of iconic gesture rate using forward entry in step 1. In step 2, total motivation and intermittent summary length were entered in the model. In Step 1 only the verbal working memory measure was included. This model was statistically significant $F\ (1, 25) = 7.267$, $p < .05$ and explained 19.4% of the variance in iconic gesture rate (see Table 20). After the forward entry of total motivation and intermittent summary length in step 2, the model explained 34.8% of the variance in iconic gesture rate [$F\ (3, 23) = 5.622$, $p < .01$]. The addition of total motivation and intermittent summary length explained an additional 19.8% of the variance in iconic gesture rate after controlling for the influence of the verbal working memory measure [$\Delta R^2 = .198$; $F\ (2, 23) = 3.944$, $p < .05$]. In the final model, both the verbal working memory ($\beta = -.598$, $p < 0.01$) and total motivation ($\beta = .470$, $p < .05$) measures were statistically significant with the verbal working memory measure having the stronger Beta value. Intermittent summary length ($\beta = -.032$, $p =$ n.s.) was not statistically significant.

Table 20

Regression model predicting iconic gesture rate

	R	R^2	*Adj.* R^2	ΔR^2	B	SE	β	t
Step 1	.475	.225	.194					
Verbal WM					-0.270	0.100	-.475	-2.696*
Step 2	.650	.423	.348	.198				
Verbal WM					-0.341	0.094	-.598	-3.634**
Total MRQ					0.504	0.185	.471	2.719**
Tokens (I)					0.000	0.002	-.032	-0.190

*$p < .05$; **$p < .01$; ***$p < .001$

Summary of results relevant to hypothesis 5: It was hypothesized that iconic gesture production would be negatively predicted by the verbal working memory measure and positively predicted by the visuospatial working memory measure. The results lend only partial support to this hypothesis. The measure of verbal working memory was a negative predictor of iconic gesture rate, and persisted as a significant predictor even when statistically controlling for the influence of total motivation and intermittent summary tokens. However, the measure of visuospatial working memory was not a significant predictor of iconic gesture rate.

Hypothesis 6: *Individuals who do not gesture spontaneously will perform better on Low Inference assessments than individuals who gesture spontaneously*

In order to determine whether individuals who did not gesture spontaneously had higher scores on Low Inference assessments than individuals who gestured spontaneously, multiple regression analyses were conducted. A dummy variable (D_G) was used in order to code the spontaneous non-gesturers (0) and the spontaneous gesturers (1). Forward entry was used for the D_G in step 1. In steps 2 and 3, since intermittent summary length and motivation differed significantly across spontaneous gesturers and non-gesturers, they were entered into the equation regardless of whether they were significant predictors of reading comprehension outcomes in order to determine whether any differences across spontaneous gesturers and non-gesturers persisted while controlling for these factors. For descriptive statistics across spontaneous gesturers and non-gesturers concerning Low Inference reading comprehension outcomes, see Table 21.

Table 21

Average (SD) Low Inference reading comprehension outcomes across spontaneous gesturers and non-gesturers

	Non-Gesturers (N=9)	Gesturers (N=18)
Fill in the blank (standard)	5.44 (2.19)	7.11 (1.57)
Accuracy for proposition inclusion (standard)	63.00 (8.16)	69.42 (8.00)

Fill-in-the-blank (standard) performance was influenced by whether individuals gestured spontaneously however, Accuracy for proposition inclusion was not.

Fill-in-the-blank (standard):

In the first step, the dummy variable was entered. This model was statistically significant F (1, 25) = 5.208, $p < 0.05$ and explained 13.9% of the variance in Fill-in-the-blank (standard) scores (see Table 22). Participants who gestured spontaneously had significantly higher Fill-in-the-blank scores (standard) than participants who did not gesture spontaneously. After intermittent summary length was entered, the model explained 20.2% of the variance in Fill-in-the-blank (standard) score [F (2, 24) = 4.290, $p < .05$]. However, the addition of intermittent summary length did not explain significantly more variance than the dummy variable alone [$\Delta R^2 = .091$; F (1, 24) = 2.963, p = n.s.] and furthermore, neither of the Beta values were significant in this model. After motivation was entered, the model explained 23.6% of the variance in Fill-in-the-blank (standard) score [F (3, 23) = 3.676, $p < .05$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .061$; F (1, 23) = 2.065, p = n.s.]. None of the Beta values were significant in this model.

Table 22

Regression model predicting Fill-in-the-blank (standard) performance among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	$Adj.\ R^2$	ΔR^2	B	SE	β	t
Step 1	.415	.172	.139					
D_G					1.637	0.717	.415	2.282*
Step 2	.513	.263	.202	.091				
D_G					0.993	0.786	.252	1.263
Tokens (I)					0.003	0.002	.343	1.721
Step 3	.569	.324	.236	.061				
D_G					0.518	0.837	.131	0.619
Tokens (I)					0.003	0.002	.305	1.551
Motivation					0.273	0.190	.285	1.437

$^*p < .05$; $^{**}p < .01$; $^{***}p < .001$

Summary of results relevant to hypothesis 6: It was hypothesized that individuals who do not gesture spontaneously would perform better on Low Inference assessments than individuals who gestured spontaneously. The results do not support this prediction. Individuals who gestured spontaneously performed significantly better than individuals who did not gesture spontaneously on the Fill-in-the-blank (standard) assessment. When statistically controlling for the influence of intermittent narrative length, this difference did not persist.

***Hypothesis** 7: Individuals who gesture spontaneously will perform better on High Inference assessments than individuals who do not gesture spontaneously*

In order to determine whether individuals who gestured spontaneously had higher scores on High Inference assessments than individuals who did not gesture spontaneously, multiple regression analyses were conducted. Forward entry was used for D_G in step 1. Intermittent summary length and motivation were entered in steps 2 and 3, respectively since these factors differed significantly across spontaneous gesturers and non-gesturers. They were entered into the equation in order to determine whether any differences across spontaneous gesturers and non-gesturers persisted when statistically controlling for these variables. For descriptive statistics concerning performance on High Inference reading comprehension outcomes across spontaneous gesturers and non-gesturers, see Table 23.

Table 23

Average (SD) for High Inference reading comprehension outcomes across spontaneous gesturers and non-gesturers

	Non-Gesturers (N=9)	Gesturers (N=18)
Main idea construction	1.67 (0.87)	2.56 (0.70)
Summary integration	2.22 (0.83)	2.78 (0.43)
Understanding	3.89 (1.45)	5.28 (1.36)
Informative	4.78 (2.44)	6.67 (2.03)
Enjoyment	4.67 (1.00)	5.06 (1.63)
Added details	1.44 (1.01)	1.56 (1.04)
Recall errors	5.78 (2.73)	3.22 (2.02)

Spontaneous gesturing influenced all High Inference reading comprehension outcomes with the exception of ratings of Enjoyment and Added details.

Main idea construction:

In the first step, the dummy variable was entered. This model was statistically significant F (1, 25) = 8.205, $p < .01$ and explained 21.7% of the variance in Main idea construction (see Table 24). After intermittent summary length was entered, the model explained 66.0% of the variance in Main idea construction score [F (2, 24) = 26.260, $p < .001$]. The addition of intermittent summary length explained significantly more variance than the dummy variable alone [$\Delta R^2 = .439$; F (1, 24) = 33.611, $p < .001$]. In this model, the Beta value for intermittent summary length was significant, but the Beta value for the dummy variable was not. After motivation was entered, the model explained 65.9% of the variance in Main idea construction score [F (3, 23) = 17.747, $p < .001$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .012$; F (1, 23) = 0.912, p = n.s.]. The only significant Beta value in this model was for intermittent summary length.

Table 24

Regression model predicting Main idea construction rating among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.497	.247	.217					
D_G					0.889	0.310	.497	2.864**
Step 2	.828	.686	.660	.439				
D_G					0.246	0.233	.138	1.060
Tokens (I)					0.003	0.001	.754	5.798***
Step 3	.836	.698	.659	.012				
D_G					0.342	0.254	.191	1.349
Tokens (I)					0.004	0.001	.771	5.863***
Motivation					-0.055	0.058	-.126	-0.955

*$p < .05$; **$p < .01$; ***$p < .001$

Summary integration:

In the first step, the dummy variable was entered. This model was statistically significant *F* (1, 25) = 5.342, $p < .05$ and explained 14.3% of the variance in Summary integration (see Table 25). After intermittent summary length was entered, the model explained 41.8% of the variance in Summary integration score [*F* (2, 24) = 10.321, $p < .01$]. The addition of intermittent summary length explained significantly more variance than the dummy variable alone [$\Delta R^2 = .286$; *F* (1, 24) = 12.782, $p < .01$]. In this model, the Beta value for intermittent summary length was significant, but the Beta value for the dummy variable was not. After motivation was entered, the model explained 39.7% of the variance in Summary integration score [*F* (3, 23) = 6.711, $p < .01$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .004$; *F* (1, 23) = 0.188, *p* = n.s.]. The only significant Beta value in this model was for intermittent summary length.

Table 25

Regression model predicting Summary integration rating among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.420	.176	.143					
D_G					0.556	0.240	.420	2.311*
Step 2	.680	.462	.418	.286				
D_G					0.171	0.225	.130	0.761
Tokens (I)					0.002	0.001	.609	3.575**
Step 3	.683	.467	.397	.004				
D_G					0.214	0.250	.162	.858
Tokens (I)					0.002	0.001	.619	3.541**
Motivation					-0.025	0.057	-.076	-0.434

*$p < .05$; **$p < .01$; ***$p < .001$

Understanding:

In the first step, the dummy variable was entered. This model was statistically significant F (1, 25) = 5.966, $p < .05$ and explained 16.0% of the variance in Understanding (see Table 26). After intermittent summary length was entered, the model explained 72.6% of the variance in Understanding score [F (2, 24) = 35.442, $p < .001$]. The addition of intermittent summary length explained significantly more variance than the dummy variable alone [$\Delta R^2 = .554$; F (1, 24) = 52.603, $p < .001$]. In this model, the Beta value for intermittent summary length was significant, but the Beta value for the dummy variable was not. After motivation was entered, the model explained 73.6% of the variance in Understanding score [F (3, 23) = 25.209, $p < .001$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .020$; F (1, 23) = 1.947, p = n.s.]. The only significant Beta value in this model was for intermittent summary length.

Table 26

Regression model predicting Understanding rating among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.439	.193	.160					
D_G					1.389	0.569	.439	2.443*
Step 2	.864	.747	.726	.554				
D_G					0.112	0.370	.035	0.302
Tokens (I)					0.007	0.001	.847	7.253***
Step 3	.876	.767	.736	.020				
D_G					-0.106	0.394	-.033	-0.268
Tokens (I)					0.007	0.001	.825	7.141***
Motivation					0.125	0.090	.162	1.395

*$p < .05$; **$p < .01$; ***$p < .001$

Informative:

In the first step, the dummy variable was entered. This model was statistically significant F (1, 25) = 4.553, $p < 0.05$ and explained 12.0% of the variance in the Informative rating (see Table 27). After intermittent summary length was entered, the model explained 66.6% of the variance in Informative rating [F (2, 24) = 26.920, $p < .001$]. The addition of intermittent summary length explained significantly more variance than the dummy variable alone [$\Delta R^2 = .538$; F (1, 24) = 41.848, $p < .001$]. In this model, the Beta value for intermittent summary length was significant, but the Beta value for the dummy variable was not. After motivation was entered, the model explained 65.4% of the variance in Informative rating [F (3, 23) = 17.382, $p < .001$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .002$; F (1, 23) = 0.169, p = n.s.]. The only significant Beta value in this model was for intermittent summary length.

Table 27

Regression model predicting Informative rating among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.392	.154	.120					
D_G					1.889	0.885	.392	2.134*
Step 2	.832	.692	.666	.538				
D_G					-0.024	0.620	-.005	-0.039
Tokens (I)					0.010	0.002	.834	6.469***
Step 3	.833	.694	.654	.002				
D_G					0.088	0.687	.018	0.127
Tokens (I)					0.010	0.002	.841	6.354***
Motivation					-0.064	0.156	-.055	-0.411

*$p < .05$; **$p < .01$; ***$p < .001$

Recall errors:

In the first step, the dummy variable was entered. This model was statistically significant F (1, 25) = 7.614, $p < 0.05$ and explained 20.3% of the variance in the Recall error rating (see Table 28). After intermittent summary length was entered, the model explained 60.3% of the variance in the Recall error rating [F (2, 24) = 20.729, $p < .001$]. The addition of intermittent summary length explained significantly more variance than the dummy variable alone [$\Delta R^2 = .400$; F (1, 24) = 26.177, $p < .001$]. In this model, the Beta value for intermittent summary length was significant, but the Beta value for the dummy variable was not. After motivation was entered, the model explained 58.8% of the variance in the Recall error rating [F (3, 23) = 13.356, $p < .001$]. The addition of motivation did not explain significantly more variance than the dummy variable and intermittent summary length combined [$\Delta R^2 = .002$; F (1, 23) = 0.124, p = n.s.]. The only significant Beta value in this model was for intermittent summary length.

Table 28

Regression model predicting the Recall error rating among spontaneous gesturers and non-gesturers in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	B	SE	β	t
Step 1	.483	.233	.203					
D_G					-2.556	0.926	-.483	-2.759*
Step 2	.796	.633	.603	.400				
D_G					-0.742	0.744	-.140	-0.998
Tokens (I)					-0.010	0.002	-.719	-5.116***
Step 3	.797	.635	.588	.002				
D_G					-0.628	0.825	-.119	-0.761
Tokens (I)					-0.010	0.002	-.713	-4.930***
Motivation					-0.066	0.187	-.051	-0.352

*$p < .05$; **$p < .01$; ***$p < .001$

Summary of results relevant to hypothesis 7: It was hypothesized that individuals who gestured spontaneously would outperform individuals who did not gesture spontaneously on High Inference assessments. Individuals who gestured spontaneously had significantly higher Main idea construction, Summary integration, Understanding, and Informative ratings; and significantly lower Recall error ratings than individuals who did not gesture spontaneously. However, when statistically controlling for the influence intermittent summary tokens, none of the differences across spontaneous gesturers and non-gesturers persisted.

Study 1 Discussion

The purpose of this study was to determine whether gesture production influences whether individuals adopt a verbal or visuospatial code in the context of a reading comprehension task. It was also of interest to investigate whether gesture production influences performance on reading comprehension assessments.

Part 1: Predictors of reading comprehension according to experimental condition

Hypothesis 1: *When gestures are encouraged and when they are free to be produced spontaneously, the visuospatial working memory measure will be the best predictor of reading comprehension*

The visuospatial code is reliant upon imagistic information (Paivio, 1986), and research suggests a possitive association between gesture production and imagery (e.g., Rimé et al., 1984). The first hypothesis was that among participants who were free to gesture or were encouraged to gesture throughout their narrative summaries, the visuospatial working memory measure would be the strongest predictor of reading comprehension outcomes.

Lending partial support to the hypothesis, the visuospatial working memory measure was a positive predictor of performance on Accuracy for proposition inclusion, Main idea construction, Summary integration, Understanding, Informative, and Enjoyment ratings, and a negative predictor of the Recall error rating among participants in the Control condition. In this condition, the visuospatial working memory and motivation measures were the strongest predictors of Low and High Inference reading comprehension outcomes.

Among participants in the Gesture condition, motivation was the only factor that predicted reading comprehension outcomes on both Low and High Inference measures. This finding does not lend support to the hypothesis, but does raise interesting questions concerning the relationship between motivation and gesture production. Firstly, does gesture production alter the way that participants engage in a narrative task? Previous researchers have shown that when the stakes of communication are high, individuals tend to have a higher gesture rate than when the stakes of communication are low (Kelly et al., 2011). It is possible that encouraging participants to gesture led them to feel as though the stakes of communication were high. This may explain why the motivation measures were the strongest predictors of reading comprehension outcomes in the Gesture condition. Unfortunately the measures used in the current study did not capture the influence of experimental condition upon task specific motivation: the reading comprehension measures in this study were obtained by having participants fill out blanks on a page and by having a rater evaluate the transcribed narratives. Future research assessing the expressiveness of participants based on the video recordings of their narratives may be useful in order to further explore this possibility. Another avenue for future research is to ask participants to rate their enjoyment of the narrative production task, and their motivation to complete the narrative production task. Though the results from the current study do not allow for an evaluation of whether task-specific motivation differed according to experimental condition, it is interesting to note that in Part 3 of the analyses, a regression predicting iconic gesture rate indicated that both motivation scores and verbal working memory are significant predictors. Motivation was a positive predictor of iconic gesture rate indicating that intrinsic motivation for reading influences how participants use their hands and arms while describing information that they have read.

It is interesting to note that in the Control condition, the visuospatial working memory and motivation measures were consistently the strongest predictors of reading comprehension outcomes. This leads to a second question: Is motivation more strongly associated with a visuospatial code than a verbal code? Some research lends support to this possibility. The vividness of mental imagery tends to be positively associated with levels of reader enjoyment

(Block, 2004; Center, Freeman, Robertson, & Outhred, 1999; Sadoski, 1983), levels of reader engagement (Langer, 1990; Long, Winograd, & Bridge, 1989; Sadoski & Quast, 1990; Schraw et al., 1995; Wade, Buxton, & Kelly, 1999) and positive emotion (Farah, 1995; Langer, 1995; Sadoski, 1999; Schraw et al., 1995; Wade et al., 1999). According to Woolley (2011), "the degree of involvement, enjoyment, and interest in reading is enhanced through the generation of suitable images" (p. 87). This is of critical importance since positive emotional responses to reading can enhance reading comprehension (Woolley, 2011).

Neurological research supports an association between the visuospatial code and affect more broadly. In a review by Mar (2003) the neurological underpinnings of narrative production and narrative comprehension were investigated. In this review, Mar (2003) argues that working memory processes are critically important for selecting and sequencing information that is being produced or comprehended. The retrieval of elaborative information that serves to enrich or add realism to a story seems to be reliant upon the posterior regions of the cingulate. Mar (2003) notes that this area may also contribute to the visuospatial imagery and could serve the function of modulating memory as a function of affect.

Smithson and Nicoladis (under review) investigated whether verbal working memory or visuospatial working memory measures predicted physiological engagement (as measured by galvanic skin response) in a narrative production task. The results of this study revealed that physiological engagement was positively associated with the measure of visuospatial working memory but not the measure of verbal working memory. This study lends support to the assertion that affect more broadly, may be more strongly associated with the visuospatial code rather than the verbal code.

Though none of this research conclusively shows that motivation is more strongly linked with a visuospatial code, it does suggest that this could be an important avenue for future research.

Hypothesis 2: *When gestures are restricted, factors associated with a verbal strategy of encoding (i.e., vocabulary and verbal working memory measures) will be the best predictors of reading comprehension*

The verbal code is thought to be superior at representing abstract knowledge in comparison to the visuospatial code (e.g., Kosslyn et al., 1995a). Previous research has shown that when gesture production is restricted, individuals tend to adopt an abstract strategy in a problem-solving task. The second hypothesis was that verbal factors (measures of vocabulary and verbal working memory) would be the best predictors of reading comprehension outcomes among individuals restricted from moving their hands. This hypothesis was largely supported since the vocabulary measure emerged as the only significant predictor of reading comprehension outcomes on both the Low and High Inference measures. More specifically, with respect to the Low Inference measures, PPVT was a positive predictor of Fill-in-the-blank (standard) performance and Accuracy for proposition inclusion. With respect to High Inference measures, PPVT was a negative predictor of the Recall error rating. It is important to note that vocabulary was only a significant predictor of one out of the seven High Inference measures. It may be the case that verbal factors are superior predictors of Low Inference assessments in comparison to High Inference assessments. Johnson-Glenberg (2000) argued that in their study a group of participants trained to use verbal strategies tended to excel at “recalling factual, linear, highly verbal, text-explicit information” (p. 780). The Low Inference assessments required participants to do just that. They had to recall information either in a verbatim manner for the Fill-in-the-blank assessment, or in a factual and text-explicit manner for the Accuracy for proposition inclusion measure.

In sum, the results of the present study suggest that when individuals are restricted from gesturing, they may rely upon a verbal code in order to complete reading comprehension tasks.

Part 1 Implications: Mixed findings in previous research may be explained by the adoption of different codes

Some researchers have found that measures of verbal working memory capacity predict reading comprehension (e.g., Cain, Oakhill, & Bryant, 2004; Swanson & Howell, 2001), whereas others have not (e.g., Dufva, Niemi, & Voeten, 2001; Oakhill et al., 1986). Similarly, some researchers have found that measures of visuospatial working memory predict reading comprehension (e.g., Goff et al., 2005; St. Clair-Thompson, 2007) whereas others have not (e.g., Seigneuric et al., 2000; Swanson & Howell, 2001).

It may be the case that the procedural differences in these studies are responsible for these conflicting results. For example, in studies that have found that verbal working memory measures predict reading comprehension, standardized tests such as the Neale Analysis of Reading Ability (e.g., Cain, 2006; Cain, Oakhill, & Bryant, 2000; Cain et al., 2004; Oakhill et al, 2003) or the Woodcock Johnson Reading Master Test (Swanson & Howell, 2001) have been used. In these assessments participants are either asked a series of comprehension questions after they read a story (Neale Analysis of Reading Ability), or to fill in blanks as they read the story (Woodcock Johnson Reading Master Test). In the studies that have not found a positive association between verbal working memory measures and reading comprehension outcomes, the reading comprehension tasks were quite different. In the study by Dufva et al. (2001), reading comprehension was assessed by having participants recount in their own words everything that they could remember from the text. Additionally, reading comprehension was assessed by asking questions about the main ideas of the reading passage. The authors did not specify whether instructions were given to children regarding their movement, and therefore, it is assumed here that they were free to move during this task. In the study by Oakhill et al. (1986), children were asked to listen to stories and they were not told that they would later be asked questions about them. In studies that have found that visuospatial working memory measures predict reading comprehension, multiple-choice questions have been used (e.g., Goff et al., 2005; St. Clair-Thompson, 2007). In studies that have found that visuospatial working memory measures do not predict reading comprehension, fill-in-the-blank questions have been used (e.g., Seigneuric et al., 2000; Swanson & Howell, 2001). In sum, different reading comprehension procedures may influence whether or not a verbal of visuospatial code is relied upon.

Part 2: Performance on Low Inference and High Inference measures of reading comprehension according to experimental condition

Hypothesis 3: *Individuals who are restricted from moving will perform better on Low Inference assessments*

Researchers believe that verbal text-based models retain verbatim information contained within a text, whereas visuospatial situation models do not (Pearson & Johnson, 1978; Stull & Mayer, 2007). In a study by Johnson-Glenberg (2000), participants were assigned to a training program that was intended to enhance either their verbal or visuospatial strategies. They found that the verbal group tended to perform better on tasks that required the recall of text-explicit information. In the current study, it was predicted that individuals would adopt a verbal code in the gesture restriction condition and therefore, they would excel on Low Inference reading comprehension tasks.

In support of this prediction, participants in the Restricted condition had significantly higher Fill-in-the-blank (standard) scores in comparison to participants in the Control condition. This difference persisted even when statistically controlling for the influence of intermittent summary tokens. Participants did not differ across conditions with respect to performance on the assessment of Accuracy for proposition inclusion. These results therefore lend only partial support to the hypothesis, however they suggest that being restricted from moving may heighten an individuals' reliance upon a verbal code, thereby conferring some benefits with respect to the completion of Low Inference reading comprehension tasks.

Hypothesis 4: *Individuals who are encouraged to gesture will perform better on High Inference assessments*

Researchers argue that the visuospatial situation model consists of elaborative inferences (Pearson & Johnson, 1978; Stull & Mayer, 2007). This suggests that individuals who adopt a visuospatial situation model may be more adept at generating inferences in comparison to individuals who rely primarily upon a verbal code. In a study by Johnson-Glenberg (2000), verbal and visuospatial training practices were assessed as interventions for reading comprehension. Participants who were trained to use visuospatial strategies tended to excel at generating implicit open-ended questions and they also excelled in creating inferences. In the present study it was hypothesized that individuals who gestured would excel on High Inference reading comprehension measures.

Minimal support for this hypothesis emerged from the analyses. Participants in the Gesture condition had significantly higher Added detail ratings than participants in the Control condition. This difference persisted when statistically controlling for the influence of intermittent summary tokens. The experimental group that performed best on the High Inference measures of reading comprehension was the Written group. Participants in the Written condition had significantly higher ratings on the Main idea construction, Summary integration, and Informative ratings in comparison to participants in the Control condition. Participants in the Written condition also had significantly lower ratings on the Recall error measure in comparison the participants in the Control condition. Though these results tend to favor participants in the written condition, it is important to note that these results may be biased by the fact that participants in the oral conditions produced a variety of verbal hesitations and dysfluencies that may have negatively influenced ratings of their narratives. Additionally, potentially important information conveyed through the gestures and facial expressions of participants in the oral conditions was not conveyed to the rater since only the transcription of the words used by the participants was used for the High Inference ratings.

Part 2 Implications: Studies where performance upon Low Inference and High Inference measures may indicate different verbal or visuospatial code reliance

The results lend weak support to Hypotheses 3 and 4. When evaluating performance on the Low Inference assessments, the only significant group difference in reading comprehension outcome favored participants in the Restricted condition. When evaluating performance on the High Inference assessments, the only significant group differences in reading comprehension outcomes favored participants in the Gesture and Written conditions. Though only weak support emerged to support the hypotheses, the results suggest that being instructed to gesture or to remain still can have an important influence on how individuals perform on Low and High inference reading comprehension measures.

A variety of researchers use methodologies to assess reading comprehension outcomes that require participants to use oral strategies (Chinn, Anderson, & Waggoner, 2001; Dennis & Moldof, 1983; McKeown, Beck, & Blake, 2009; Nystrand, 1997; Saunders & Goldenberg, 1999). Gesture production is likely to occur in these cases and may affect performance on High Inference reading comprehension outcomes. Other researchers use methodologies to assess reading comprehension outcomes that require individuals to remain still as they complete the task (e.g., Berl et al., 2010; Humphreys & Gennari, 2014; Landi, Frost, Mencl, Sandak, & Pugh, 2013; Moss, Schunn, Schneider, McNamara, & VanLehn, 2011). The results of the current study suggest that these participants may excel on Low Inference reading comprehension tasks.

Part 3: Performance on Low Inference and High Inference measures of reading comprehension according to whether individuals gestured spontaneously or not

In the Control condition, 18 participants gestured spontaneously while 9 participants did not. These participants did not differ on any of the independent measures with the exception of motivation. Individuals who gestured spontaneously tended to have higher ratings of motivation. This is an important consideration since it suggests that gesture production that occurs spontaneously in a reading comprehension task may be a useful index of intrinsic motivation for reading. This finding fits well with previous research by Kelly et al. (2011), showing that individuals had higher gesture rates when the stakes of communication were high as opposed to when they are low.

Intermittent summary length also differed significantly across individuals who gestured spontaneously and those who did not gesture spontaneously. Since both Motivation and intermittent summary length differed significantly across spontaneous gesturers and non-gesturers these factors were statistically controlled in all analyses in this section.

Hypothesis 5: *Iconic gesture production is negatively associated with the measure of verbal working memory and positively associated with the measure of visuospatial working memory*

Previous researchers have suggested that individuals with strong spatial skills and low verbal skills tend to have high gesture rates (Hostetter & Alibali, 2007). It was predicted that iconic gesture production would be positively predicted by the visuospatial working memory measure and negatively predicted by the verbal working memory measure. Lending support to this hypothesis, the results revealed that the verbal working memory measure was a significant negative predictor of iconic gesture rate. Furthermore, the verbal working memory measure persisted as a negative predictor of iconic gesture production even when statistically controlling for both motivation and intermittent summary tokens. Contrary to the hypothesis, the visuospatial working memory measure did not predict iconic gesture production. These results suggest that individuals who have difficulty storing and manipulating language information may have used iconic gestures in a compensatory manner as they relayed their narratives about the text that they had read.

Hypothesis 6: *Individuals who do not gesture spontaneously will perform better on Low Inference assessments than individuals who gesture spontaneously*

Just as it was predicted that individuals in the Restricted condition would demonstrate advantages on Low Inference measures, it was also predicted that the spontaneous non-gesturers would demonstrate the same advantages. Johnson-Glenberg (2000) found that individuals who received verbal reading comprehension training tended to excel on measures of reading comprehension that required the participants to recall text-explicit information. It was predicted that spontaneous non-gesturers would be more strongly reliant upon a verbal code and would therefore show advantages on Low Inference assessments of reading comprehension.

Participants who spontaneously gestured had significantly higher Fill-in-the-blank (standard) scores in comparison to participants who did not gesture spontaneously. When the number of intermittent summary tokens was entered into the regression, the difference between spontaneous gesturers and non-gesturers was no longer present. Participants who spontaneously gestured did not differ from those who did not gesture on the measure of Accuracy for proposition inclusion.

The results are in direct opposition to what was hypothesized. Participants who gestured spontaneously rather than those who did not gesture spontaneously, showed an advantage on the Fill-in-the-blank (standard) assessment. When controlling for intermittent summary length, this difference was no longer significant. This suggests that by spontaneously gesturing, participants were able to increase the length of their summaries and this in turn contributed to a superior ability to recall specific words that were used in the reading passage.

***Hypothesis** 7: Individuals who gesture spontaneously will perform better on High Inference assessments than individuals who do not gesture spontaneously*

Just as it was predicted that individuals in the Gesture condition would show advantages on High Inference measures, it was also predicted that participants who spontaneously gestured in the Control condition would demonstrate similar advantages. Johnson-Glenberg (2000) found that individuals who received visuospatial reading comprehension training tended to excel on recalling text-implicit information. It was predicted that individuals who gestured spontaneously would be strongly reliant upon a visuospatial code and as a result would outperform the non-gesturers on the High Inference reading comprehension measures.

The results lend support to this prediction since participants who gestured spontaneously outperformed the non-gesturers on every High Inference measure of reading comprehension with the exception of the Enjoyment and Added details ratings. All of these differences became non-significant when statistically controlling for intermittent summary tokens. This suggests that the reading comprehension advantages seen among the gesturers may be due to the length of their narrative summaries rather than gesture production per se.

Study 1 Conclusion

Collectively, the results from the current study suggest that individuals who are free to gesture may rely on both visuospatial and motivational resources to complete a reading comprehension task; individuals who are encouraged to gesture may rely on motivational resources to complete a reading comprehension task; and individuals who are restricted from gesturing may rely on verbal resources in order to complete a reading comprehension task. The fact that gestures may influence the reliance upon motivational and working memory resources is critically important for researchers who investigate reading comprehension. When reporting results, it would be useful for researchers to provide explicit details concerning participant movement during reading comprehension tasks.

The results from the current study also suggest that spontaneous gesturing may be useful for the completion of a reading comprehension task. Spontaneous gesturers had superior performance on both Low and High Inference reading comprehension assessments in comparison to non-spontaneous gesturers. When intermittent summary tokens were statistically controlled, all of the beneficial effects of spontaneous gesturing disappeared. This suggests that it was the production of longer narratives that was beneficial for reading comprehension among the spontaneous gesturers.

Two important questions stem from the findings generated in Study 1:

(1) Why does spontaneous gesturing lead to advantages on Low Inference measures of reading comprehension whereas encouraged gesture production does not?

Gesturing spontaneously as opposed to being encouraged to gesture, is associated with advantages on Low Inference measures of reading comprehension. One explanation for this finding concerns the negative correlation between gesture rate and the measure of verbal working memory. Individuals who had a high propensity to gesture tended to be the ones who struggled to store and manipulate language information. Individuals with weak verbal working memory resources may use gesture production as a compensatory mechanism. In support of this argument, research suggests that gestures may play an important role in accessing words and linguistic constructions (Alibali, Kita & Young, 2000; Frick-Horbury & Guttentag, 1998; Krauss, Chen, & Gottesman, 2000; Morrel-Samuels & Krauss, 1992; Rauscher, et al., 1996). Importantly, lexical retrieval tends to be facilitated by the use of gestures among both children (Pine, Bird, & Kirk, 2007) and adults (Frick-Horbury & Guttentag, 1998). For example, children with language impairments tend to use a higher rate of gestures than typically developing children (Iverson & Braddock, 2011). Gesture rates tend to be higher during extemporaneous speech than during rehearsed speech (Chawla & Krauss, 1994). Gesture rates also tend to be higher during fast speech (Rauscher et al., 1996). Previous research has shown that iconic gesture rate is negatively associated with verbal short-term memory, suggesting that iconic gestures may serve a compensatory relationship when individuals have difficulties storing information in a verbal code (Smithson & Nicoladis, 2013).

(2) How does the association between gesture production and narrative length differ among the spontaneous gesturers and the encouraged gesturers?

The narrative lengths of participants in the Gesture, Control, and Restricted conditions were comparable. In contrast, the narrative lengths of spontaneous gesturers were significantly longer than the narrative lengths of spontaneous non-gesturers. It is important to note that in the Control condition, iconic gesture production was negatively predicted by verbal working memory. It may be the case that since spontaneous gesturers struggled to use the verbal code they relied more strongly upon the use of a visuospatial code and this may have contributed to the generation of longer summaries.

Research has accumulated to suggest that the reliance upon a visuospatial code results in longer narrative productions in comparison to the reliance upon a verbal code. This has been revealed by presenting participants with either a verbal or visuospatial representation of the same information and asking participants to relay a narrative about what they can remember from the information that they were presented with. In a study by Hostetter and Skirving (2011), participants were either assigned to hear a verbal description of a cartoon two times (verbal condition), or hear a verbal description of a cartoon once and watch a cartoon (the cartoon that was described) (visuospatial condition). Participants in the visuospatial condition produced verbal descriptions of the cartoons that were marginally longer than participants in the verbal condition. Additionally, participants in the visuospatial condition described significantly more events per story than participants in the verbal condition. Similarly, in a study by Parrill, Bullen, and Hoburg (2010), participants were either assigned to watch a cartoon clip and describe it to a listener (visuospatial condition) or read three text versions of the same cartoon stimuli (verbal condition). Participants then described what they could remember to a listener. Participants in the verbal condition used significantly fewer utterances. Together, these results suggest that narrative productions that rely upon a visuospatial code may be longer than narrative productions that rely upon a verbal code.

A similar effect is seen when varying speech topic. For example, in a study by Feyereisen and Havard (1999), young and old participants were asked to describe abstract, visual, and motor information to a listener. The results revealed that participants' verbal responses were shorter when discussing abstract topics in comparison to visual information. Note however that this was only significant among the older adults.

In sum, it may be the case that individuals with a tendency to spontaneously gesture had weak verbal working memory capacities and therefore relied primarily upon a visuospatial code in order to complete the reading comprehension task. The reliance upon the visuospatial code may have resulted in longer participant narratives and this may explain the reading comprehension advantages seen among the spontaneous gesturers on both Low and High Inference reading comprehension outcomes. More specifically, by using more words in their narrative summaries, it is possible that spontaneous gesturers more robustly encoded the information that was read in the text.

It should be noted that this is a very tentative explanation for the findings since all that can be concluded from the data is that spontaneous gesturers tend to tell longer narratives and the act of producing longer narratives seems to confer advantages on both Low and High Inference tasks. Furthermore, due to a small sample size (particularly for the analyses conducted in Part 3), one limitation of the current study is statistical power. Future research is required in order to assess whether these patterns of results would be replicated among larger samples.

One perplexing result that emerged from this study was that encouraging gesture production only resulted in a higher rating of Added details in comparison to participants in the Control condition. Spontaneous gesturers demonstrated advantages on the majority of the reading comprehension measures. Since spontaneous gesture production was negatively associated with the verbal working memory measure, it was hypothesized that gesture production may be particularly useful for individuals who struggle to maintain and manipulate verbal information mentally. There are a number of ways in which verbal working memory could be taxed in a reading comprehension task. Verbal working memory may be especially taxed in a reading comprehension task among children since verbal working memory capacity increases across development (Gathercole et al., 2004). This possibility was addressed in Study 2.

CHAPTER III

Introduction to Study 2

The results from Study 1 suggest that individuals who experience a strain on their verbal working memory resources use gesture production spontaneously and this supports the generation of longer and more accurate summaries, and leads to superior reading comprehension outcomes in comparison to individuals who do not gesture spontaneously. Encouraging gesture production did not enhance reading comprehension outcomes in Study 1, however encouraging gesture production may be beneficial among participants whose verbal working memory resources are strained. In order to investigate the tenability of this prediction, it was necessary to assess the usefulness of encouraging gesture production among a population where the demands upon verbal working memory resources are high. One population whose verbal working memory resources may be particularly taxed during a reading comprehension task is children. Though the structure of working memory with the three distinct components of verbal storage, visuospatial storage, and central executive are present from the age of 6 years, these three components continue to increase in functional capacity across the school years to adolescence (Gathercole et al., 2004). Across development, individuals are better able to use flexible strategies and processing resources in order to bolster the storage capacities of verbal working memory and visuospatial working memory (Gathercole, Pickering, Ambridge, & Wearing, 2004).

Children were therefore used in the current study in order to test the possibility that encouraging gesture production could be beneficial among participants whose verbal working memory resources are strained. To increase the demands upon verbal working memory resources to an even greater degree, three reading passages of increasing difficulty were used among children. These stories were longer and the vocabulary was more difficult with each increasing level.

The transition from “learning to read” to “reading to learn” is thought to occur throughout the fourth grade when children are approximately 8 or 9 years of age (Chall, 1983). It was of interest to assess reading comprehension among children who had transitioned to the “reading to learn” stage (to increase task comparability with adults). It was also of interest to assess reading comprehension among children whose verbal working memory capacity was as different from the adults as possible. Grade 5 children (ranging in age from 9-11 years) were therefore included in this study. Three research questions were investigated:

Research question 1: *Does gesture production influence reading comprehension outcomes on Low Inference and High Inference assessments?*

It was predicted that participants in the Gesture condition would have higher reading comprehension scores in comparison to participants in the Control condition. It was predicted that with an increase in reading passage difficulty, the usefulness of gesture production would become more apparent. Therefore, the strongest advantages were expected for participants in the Gesture condition when they completed the most difficult reading comprehension passage.

Research question 2: *Is narrative length affected by experimental condition?*

In Study 1, individuals who gestured spontaneously told significantly longer narratives than individuals who did not gesture spontaneously. It was hypothesized that participants in the Gesture condition would relay longer narratives and that the difference in narrative length would increase between the Control and Gesture conditions as reading passage difficulty increased.

Research question 3: *Does statistically controlling for intermittent summary length influence the strength of any differences between the Control and Gesture groups on measures of Low Inference and High Inference recall on a standardized reading comprehension assessment?*

In Study 1, the advantages among individuals who gestured spontaneously did not persist when statistically controlling for intermittent summary length. It was argued that gesture production is beneficial for reading comprehension outcomes because it leads to the production of longer narratives. In the current study it was predicted that any gestural advantages on Low Inference and High Inference reading comprehension measures would disappear when statistically controlling for intermittent summary length.

Method

Participants

Twenty-one grade five participants were recruited from elementary schools in Edmonton, Alberta. The participants ranged in age from 9 to 11 years (M = 10.19, SD = 0.51) with 13 males and 8 females. Participants were randomly assigned to one of two experimental conditions for the summary production component of the study: (1) *Gesture condition*: Participants in this condition were encouraged to use meaningful hand and arm movements as they relayed their narratives, (2) *Control condition*: Participants were not provided with any instructions regarding their movements in this condition. Outliers were identified as having values with $Z > \pm 3.29$ (Tabachnick & Fidell, 2007) and the individual values were removed from analysis. One participant had a vocabulary score that was an outlier, and one participant had both working memory scores that were outliers. In terms of missing data, one participant was missing a vocabulary score (due to time constraints), and final summaries were not obtained from one participant for the Level 3 assessment, four participants for the level 4, and two participants for the level 5 assessment (this occurred in cases where the experimenters forgot to ask the children to provide a final summary in addition to the intermittent summaries).

Reading passages for children. The reading passages for children were taken directly from the Neale Analysis of Reading Ability (NARA) assessment (Neale, 1999). In this assessment, there is a practice reading followed by reading passages that correspond to 6 levels of difficulty. All participants completed the practice reading. With respect to the test reading passages, children started at the 3rd level of difficulty (following the recommendations of the manual) (Neale, 1999) and also completed the 4th and 5th levels. All of the narratives were presented on paper.

Each of the reading passages had a speaking head icon after every main paragraph. When participants arrived at this speaking head icon they were asked to provide a summary of the paragraph just before the icon (and only a summary of that paragraph). For level 3 and 4 reading passages there were two speaking head icons and for the level 5 reading passage there were three speaking head icons throughout the text. The different number of speaking head icons was used since the narratives increase in length as the difficulty of the reading passages increases. At the end of each reading passage, participants were asked to answer reading comprehension questions (6 for the practice passage and 8 for the test passages) and to provide a comprehensive summary of everything they could remember from the entire story.

The materials used for the vocabulary assessment, Motivations for Reading Questionnaire and the working memory assessments were identical to the materials used with adults. With respect to the working memory assessment, only visuospatial working memory task (Odd-One-Out) and Verbal working memory task (Listening Recall) were measured among children. These are the same tasks that were used by adults (see Study 1).

Procedure. Informed consent was obtained from parents of all children who participated in this study and assent was obtained from all children participating in this study. Children were asked to complete the following items throughout two sessions:

1) AWMA (Session 1): Participants completed the working memory assessments.

2) *Reading comprehension task (Session 2):* Children initially completed a practice reading passage from the NARA assessment (Neale, 1999). They were asked to stop reading when they came to a speaking head icon. At this point children in the Gesture condition were asked: “Can you tell me everything that you can remember from that paragraph while trying to move your hands and arms in meaningful ways as you tell me?” and children in the Control condition were asked “Can you tell me everything that you can remember from that paragraph?”. Once

participants got the end of the practice reading passage they were asked 6 reading comprehension questions. Subsequently, they were asked to relay comprehensive summaries of the entire reading passage. Children in the Gesture condition were asked: “Can you tell me everything that you can remember from that entire story while trying to move your hands and arms in meaningful ways as you tell me?” and children in the Control condition were asked “Can you tell me everything that you can remember from that entire story?”. The same procedure was followed for reading passages of increasing difficulty with the only exception being that 8 open ended reading comprehension questions rather than 6 were asked.

3) *PPVT (Session 2):* Participants completed the vocabulary assessment. The experimenter said a word and the participant was shown four black and white pictures. The participant was asked to point to the picture that best corresponded to the word that the experimenter said.

4) *MRQ (Session 2):* Participants answered 20 questions in Likert scale format regarding their intrinsic motivations to read.

Speech and gesture coding, narrative evaluations, and decoding speeds were obtained in the same manner as with adults (see Study 1) (note that Accuracy for proposition inclusion was not assessed in Study 2). Reading comprehension questions were labeled by a trained coder as ‘Low Inference’ or ‘High Inference’. A trained coder provided a score for participant responses to the NARA questions on each level of assessment.

Results

Descriptive statistics for age, decoding speed, PPVT, MRQ, verbal working memory, and visuospatial working memory can be found in Table 29. T-tests indicate that none of these variables differed significantly according to experimental condition.

Table 29

Average (SD) independent variable measures according to experimental condition

Measure	Range	Control (N=11)	Gesture (N=10)	*t*-test
Age	9 - 11	10.18 (0.40)	10.20 (0.63)	*t* (19) = -0.079, *p* = n.s., *Cohen's d* = -0.038
Gender		6 boys; 5 girls	7 boys 3 girls	
Decoding speed (Level 3)	25 - 58	37.82 (9.31)	33.80 (6.00)	*t* (19) = 1.161, *p* = n.s., *Cohen's d* = 0.513
Decoding speed (Level 4)	32 - 88	54.55 (14.70)	46.20 (12.61)	*t* (19) = 1.389, *p* = n.s., *Cohen's d* = 0.610
Decoding speed (Level 5)	47 - 101	72.30 (17.06)	64.70 (15.76)	*t* (18) = 1.035, *p* = n.s., *Cohen's d* = 0.463

		Control (N=10)	Gesture (N=9)	
PPVT	95 - 142	115.20 (12.73)	122.56 (12.32)	*t* (17) = -1.277, *p* = n.s., *Cohen's d* = -0.588
Total MRQ	9.37 – 15.80	12.94 (2.13)	13.19 (1.54)	*t* (19) = -0.305, *p* = n.s., *Cohen's d* = -0.135
		Control (N=10)	Gesture (N=10)	
Verbal WM	79 - 137	112.28 (18.91)	103.61 (13.13)	*t* (18) = 1.191, *p* = n.s., *Cohen's d* = 0.533
Visuo WM	93 - 137	110.93 (15.43)	103.74 (10.93)	*t* (18) = 1.202, *p* = n.s., *Cohen's d* = 0.538

Research question 1: *Does gesture production influence reading comprehension outcomes on Low Inference and High Inference assessments?*

Analyses were conducted in order to assess the influence of gesture production on reading comprehension outcomes on the NARA assessment and on reading comprehension ratings of the final summaries produced by participants.

NARA assessment

For descriptive statistics concerning participants' performance on the NARA assessment across conditions and levels of reading passage difficulty, see Table 30.

Table 30

Average (SD) NARA percentages according to experimental condition

Type		Control (N=11)	Gesture (N=10)
Low Inference	Level 3	47.73 (24.41)	63.13 (31.52)
	Level 4	53.79 (26.32)	60.42 (18.66)
	Level 5	28.79 (23.68)	57.08 (28.19)
High Inference	Level 3	43.18 (16.17)	56.25 (21.45)
	Level 4	54.55 (36.77)	52.50 (18.45)
	Level 5	43.18 (29.77)	72.50 (27.51)

With respect to performance on the NARA assessment, a 2x2x3 mixed factorial ANOVA was used to assess the influence of condition (a between-subjects factor) upon performance on Low Inference and High Inference reading comprehension questions (a within-subjects factor) on three different levels of assessment (a within-subjects factor).

There was a significant main effect of condition, F (1, 19) = 6.229, $p < .05$, $\eta_p^2 = 0.247$. Participants in the Gesture condition (M = 60.31, SE = 4.38) had higher scores than participants in the Control condition (M = 45.20, SE = 4.18). The main effect of question type (Low Inference or High Inference) did not reach significance, F (1, 19) = 0.187, p = n.s., $\eta_p^2 = 0.010$. Therefore, the children performed similarly upon the Low Inference and High Inference assessments of reading comprehension. The main effect of Level of assessment also did not reach significance, F (2, 38) = 0.433, p = n.s., $\eta_p^2 = 0.022$. Children performed similarly upon each level of assessment.

The interaction between Condition and question type did not reach significance, F (1, 19) = 0.148, p = n.s., $\eta_p^2 = 0.008$. The interaction between Condition and Level was not significant, F (2, 38) = 3.137, p = n.s., $\eta_p^2 = 0.142$. The interaction between Question type and Level did not reach significance, F (2, 38) = 2.403, p = n.s., $\eta_p^2 = 0.112$. The interaction between Condition, Level, and Question type also did not reach significance, F (2, 38) = 0.113, p = n.s., $\eta_p^2 = 0.006$.

See Figures 1 and 2 for Low Inference and High Inference assessment performance, respectively.

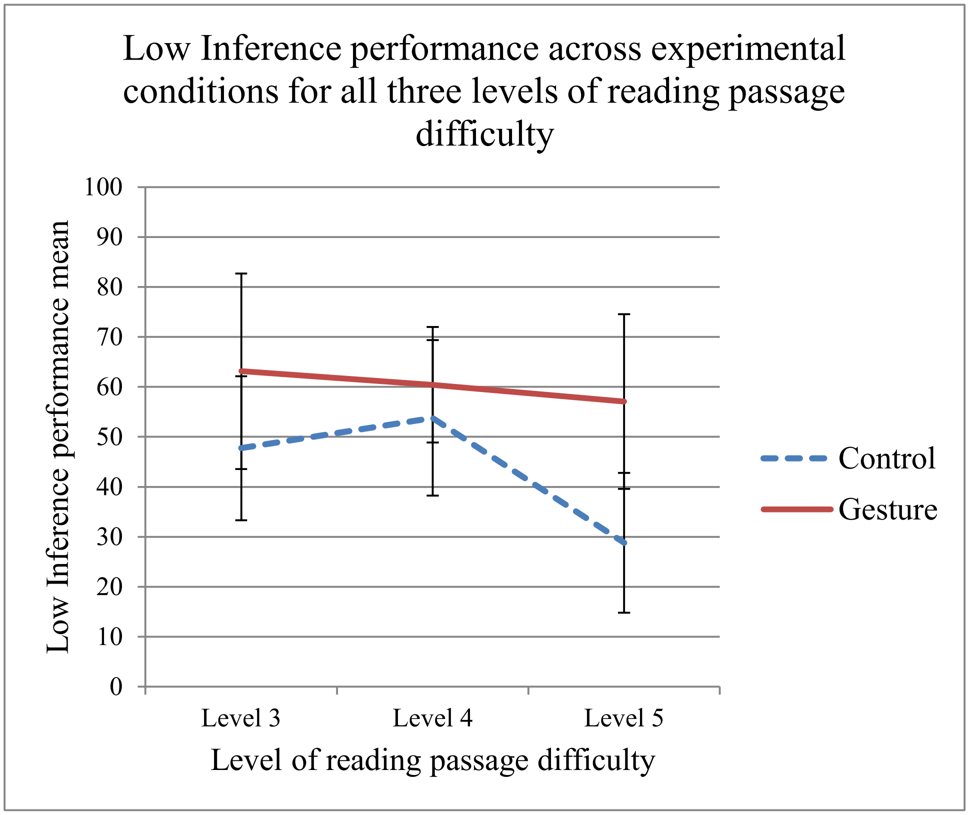

Figure 1. Performance on Low Inference NARA questions across conditions and levels of assessment. Error bars represent 95% confidence interval.

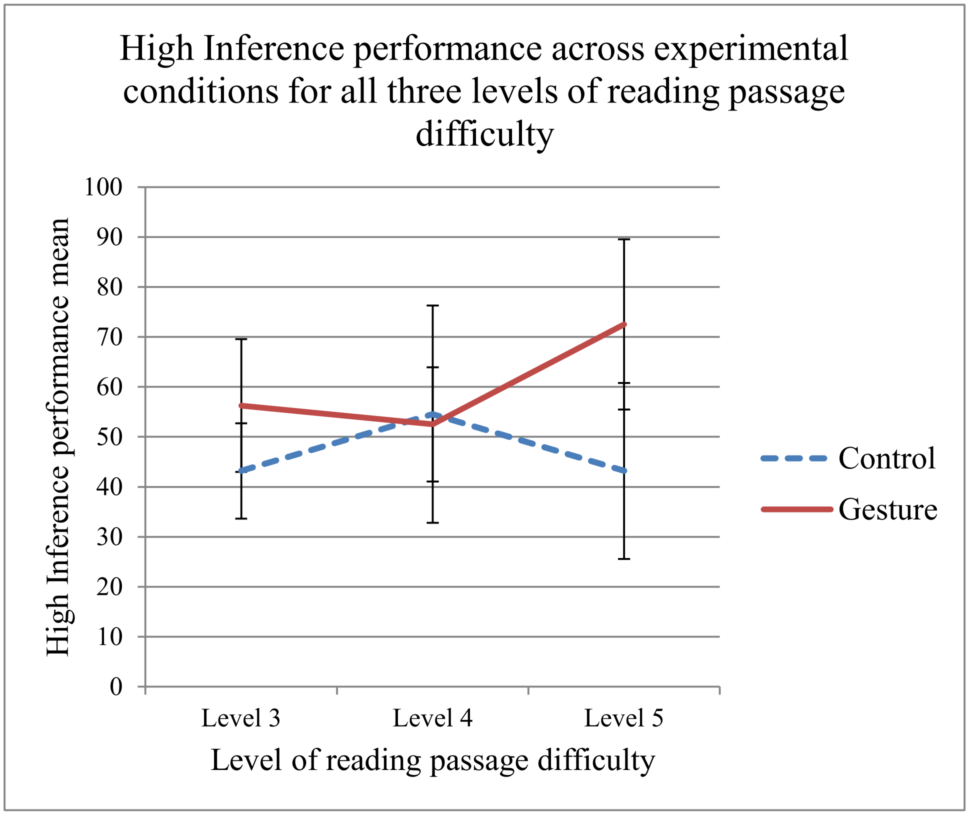

Figure 2. Performance on High Inference NARA questions across conditions and levels of assessment. Error bars represent 95% confidence interval.

Reading comprehension ratings

Ratings of the final narrative summaries were assessed across conditions. Descriptive statistics for all ratings for each level of assessment among participants in the Control and Gesture conditions can be found in Table 31.

Table 31

Average (SD) reading comprehension ratings according to experimental condition

	Level 3	
	Control (N = 11)	Gesture (N = 9)
Main idea construction	2.27 (0.79)	2.67 (0.50)
Summary integration	2.45 (0.69)	2.89 (0.33)
Understanding	4.36 (1.57)	5.22 (1.09)
Informative	5.91 (1.92)	7.44 (0.88)
Enjoyment	5.36 (1.57)	6.00 (1.00)
Added details	1.73 (1.42)	1.78 (1.72)
Recall errors	3.91 (2.34)	2.44 (1.33)
	Level 4	
	Control (N = 10)	Gesture (N = 7)
Main idea construction	2.70 (0.48)	3.00 (0.00)
Summary integration	2.60 (0.70)	2.86 (0.38)
Understanding	5.10 (0.88)	5.57 (0.79)
Informative	6.90 (1.37)	7.43 (0.98)

Enjoyment	5.80 (1.55)	6.57 (0.79)
Added details	0.90 (0.74)	1.00 (1.00)
Recall errors	3.20 (1.55)	1.86 (1.35)
	Level 5	
	Control	Gesture
	(N = 11)	(N = 8)
Main idea construction	2.00 (0.63)	2.88 (0.35)
Summary integration	2.55 (0.52)	3.00 (0.00)
Understanding	4.00 (1.26)	6.13 (0.83)
Informative	5.55 (1.57)	7.38 (1.19)
Enjoyment	5.00 (1.55)	6.75 (1.39)
Added details	3.09 (2.02)	0.75 (0.71)
Recall errors	5.18 (2.44)	2.00 (1.31)

A 2x3 mixed factorial ANOVA was used to assess the influence of condition (a between-subjects factor) and the three different levels of assessment (a within-subjects factor) on each of the narrative ratings.

Main idea construction

With respect to Main idea construction, there was a significant main effect of condition, F (1, 14) = 4.594, p = .05, η_p^2 = 0.247. Participants in the Gesture condition (M = 2.83, SE = 0.17) had higher scores than participants in the Control condition (M = 2.37, SE = 0.13). The main effect of Level of assessment was also significant, F (2, 28) = 4.061, p < .05, η_p^2 = 0.225. Approximate pairwise Bonferroni-correct post-hoc t-tests indicated that participants had higher scores on the Level 4 assessment (M = 2.85, SE = 0.10) than on the Level 5 assessment (M = 2.42, SE = 0.15). The interaction between Condition and Level was not significant, F (2, 28) = 2.042, p = n.s., η_p^2 = 0.127 (see Figure 3).

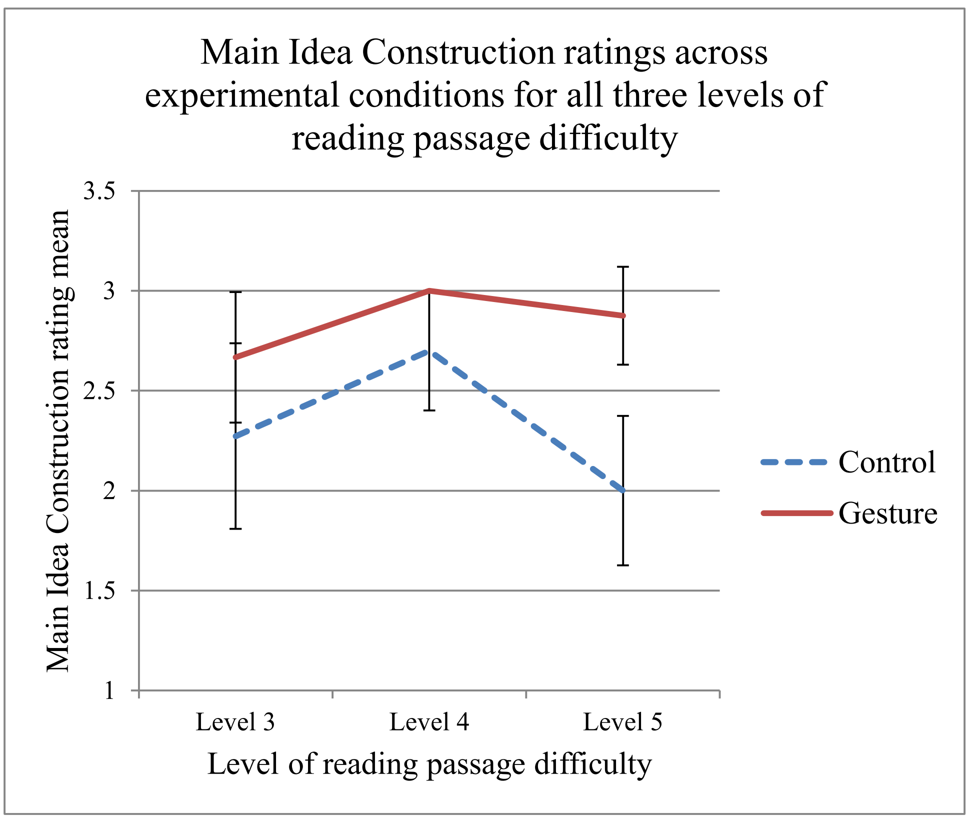

Figure 3. Performance on Main idea construction ratings across conditions and levels of assessment. Error bars represent 95% confidence interval.

Summary integration

With respect to Summary integration, the main effect of condition was not significant, F (1, 14) = 2.743, p = n.s., $\eta_p^2 = 0.164$. The main effect of Level of assessment was not significant, F (2, 28) = 0.302, p = n.s., $\eta_p^2 = 0.021$. The interaction between Condition and Level was also not significant, F (2, 28) = 0.302, p = n.s., $\eta_p^2 = 0.021$.

Understanding

With respect to Understanding, the main effect of condition was significant, F (1, 14) = 4.963, $p < .05$, $\eta_p^2 = 0.262$. Participants in the Gesture condition (M = 5.56, SE = 0.34) had significantly higher Understanding scores than participants in the Control condition (M = 4.60, SE = 0.26). The main effect of Level of assessment was not significant, F (2, 28) = 1.831, p = n.s., $\eta_p^2 = 0.116$. The interaction between Condition and Level was significant, F (2, 28) = 5.057, $p < .05$, $\eta_p^2 = 0.265$ (see Figure 4).

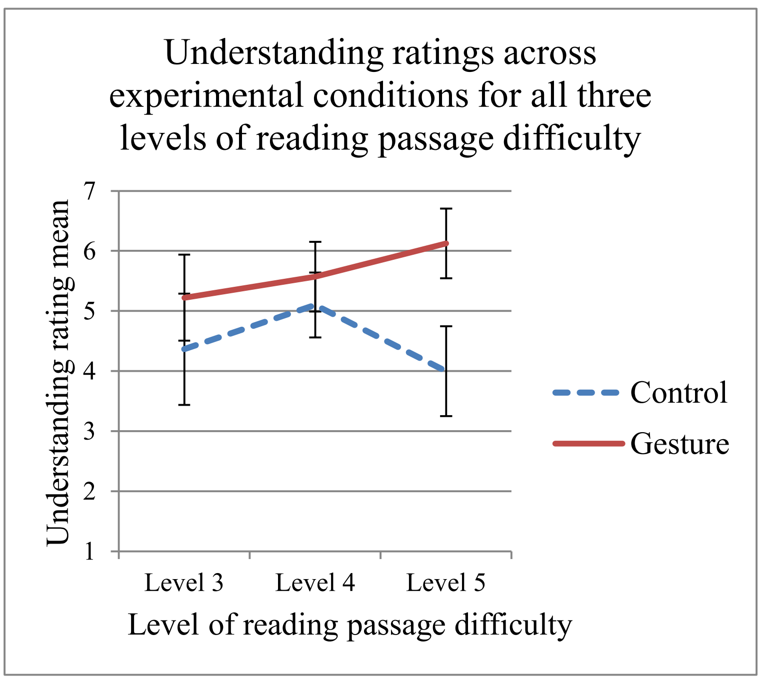

Figure 4. Performance on Understanding ratings across conditions and levels of assessment. Error bars represent 95% confidence interval.

The effect of condition was assessed at each Level of reading comprehension using independent samples t-tests (using the approximate Bonferroni-correct $\alpha/3$ in order to correct for multiple comparisons, $\alpha = 0.017$) (see Table 32). These tests revealed that participants in the Control and Gesture conditions performed comparably on the Level 3 and 4 ratings of Understanding, but participants in the Gesture condition had significantly higher Understanding ratings than participants in the Control condition on the Level 5 assessment.

Table 32

T-tests comparing Understanding rating across conditions on all levels of assessment

Level	**Control**	**Gesture**	***t*-test**
Level 3	4.36 (1.57)	5.22 (1.09)	t (18) = -1.388, p = n.s., *Cohen's d* = -0.636
Level 4	5.10 (0.88)	5.57 (0.79)	t (15) = -1.137, p = n.s., *Cohen's d* = -0.562
Level 5	4.00 (1.26)	6.13 (0.83)	t (17) = -4.127, $p < .01$, *Cohen's d* = -1.996

Informative

With respect to the Informative rating, the main effect of condition was significant, F (1, 14) = 5.803, $p < .05$, $\eta_p^2 = 0.293$. Participants in the Gesture condition (M = 7.44, SE = 0.40) had significantly higher Informative ratings than participants in the Control condition (M = 6.23, SE = 0.31). The main effect of Level of assessment was not significant, F (2, 28) = 2.494, p = n.s., η_p^2 = 0.151. The interaction between Condition and Level was also not significant, F (2, 28) = 1.021, p = n.s., η_p^2 = 0.068 (see Figure 5).

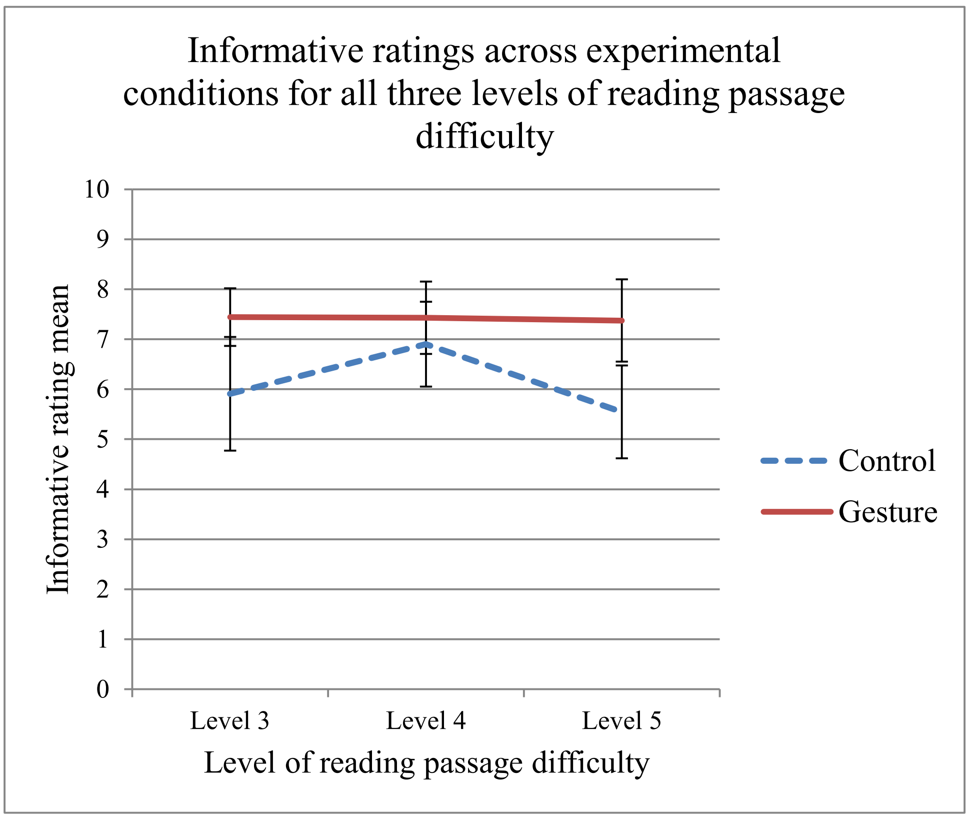

Figure 5. Performance on Informative ratings across conditions and levels of assessment. Error bars represent 95% confidence interval.

Enjoyment

When assessing the Enjoyment rating, the main effect of condition [F (1, 14) = 3.851, p = n.s., $\eta_p^2 = 0.216$], the main effect of Level of assessment [F (2, 28) = 1.013, p = n.s., $\eta_p^2 = 0.067$], and the interaction between Condition and Level [F (2, 28) = 2.079, p = n.s., $\eta_p^2 = 0.129$] were not significant.

Added details

With respect to Added details, the main effect of condition was not significant, F (1, 14) = 1.225, p = n.s., $\eta_p^2 = 0.080$. The main effect of Level of assessment was not significant, F (2, 28) = 2.205, p = n.s., $\eta_p^2 = 0.136$. The interaction between Condition and Level was significant, F (2, 28) = 3.646, $p < .05$, $\eta_p^2 = 0.207$ (see Figure 6).

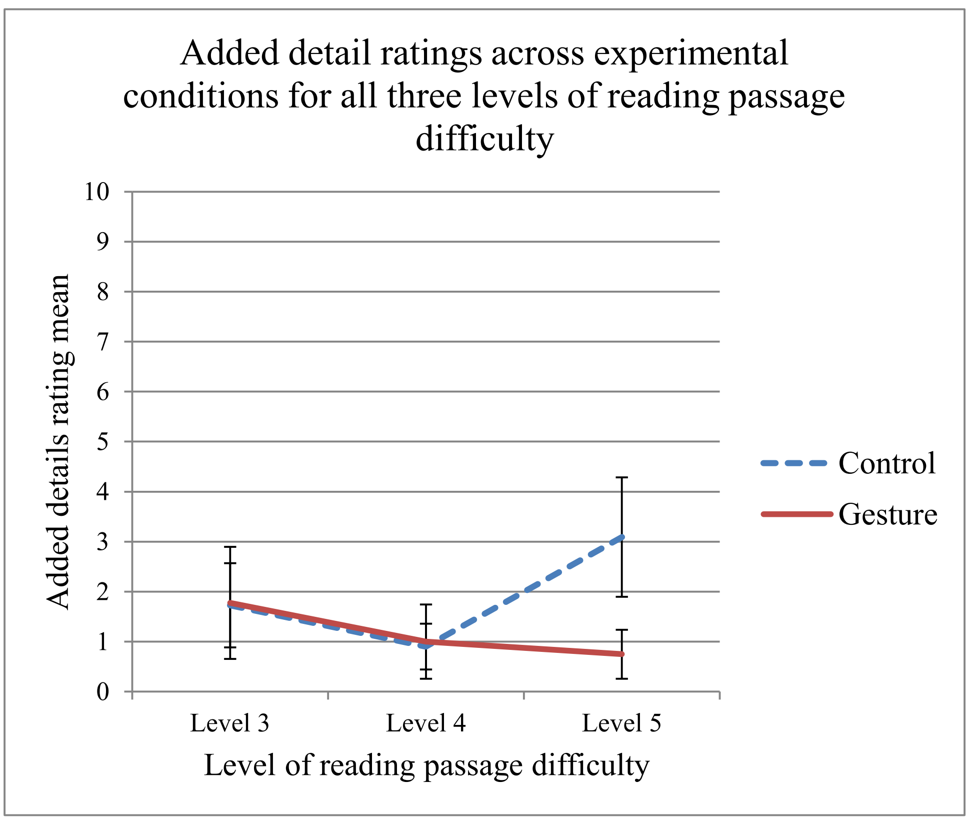

Figure 6. Performance on Added details ratings across conditions and levels of assessment. Error bars represent 95% confidence interval.

The effect of condition was assessed at each Level of reading comprehension using independent samples t-tests (using the approximate Bonferroni-correct $\alpha/3$ in order to correct for multiple comparisons, $\alpha = 0.017$) (see Table 33). The results of the t-tests indicate that the rating of Added details was comparable among participants in the Control and Gesture condition for the Level 3 and 4 assessments, however participants in the Control condition had a significantly higher rating than participants in the Gesture condition on the Level 5 assessment.

Table 33

T-tests comparing average (SD) Added details rating across conditions on all levels of assessment

Level	**Control**	**Gesture**	***t*-test**
Level 3	1.73 (1.42)	1.78 (1.72)	t (18) = -0.072, p = n.s., *Cohen's d* = -0.032
Level 4	0.90 (0.74)	1.00 (1.00)	t (15) = -0.238, p = n.s., *Cohen's d* = -0.114
Level 5	3.09 (2.02)	0.75 (0.71)	t (13.114) = 3.552, $p < .01$, *Cohen's d* = 1.546

Recall errors

With respect to the Recall error rating, the main effect of condition was significant, F (1, 14) = 6.681, $p < .05$, $\eta_p^2 = 0.323$. Participants in the Control condition ($M = 3.97$, $SE = 0.40$) had significantly higher Recall error ratings than participants in the Gesture condition ($M = 2.28$, $SE = 0.52$). The main effect of Level of assessment was not significant, F (2, 28) = 3.116, p = n.s., $\eta_p^2 = 0.182$. The interaction between Condition and Level was also not significant, F (2, 28) = 2.122, p = n.s., $\eta_p^2 = 0.132$ (see Figure 7).

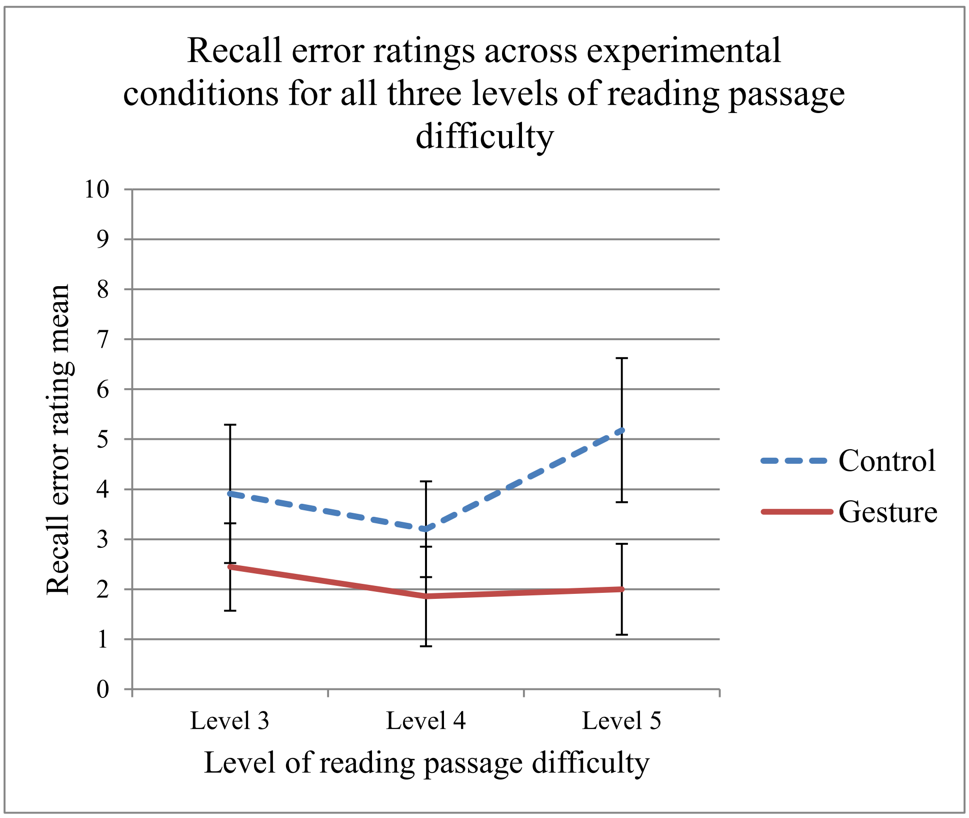

Figure 7. Performance on Recall error rating across conditions and levels of assessment. Error bars represent 95% confidence interval.

Summary of results relevant to research question 1: Does gesture production influence reading comprehension outcomes on Low Inference and High Inference assessments? On the NARA assessment, there was a significant main effect of condition in which participants in the Gesture condition had significantly higher Low and High Inference scores in comparison to participants in the Control condition. When considering the ratings of the final summaries, participants in the Gesture condition had higher Main idea construction and Informative ratings in comparison to participants in the Control condition. Furthermore, participants in the Gesture condition had lower Recall error ratings than participants in the Control condition. Participants in the Gesture condition had higher Understanding scores and lower Added detail scores than participants in the Control condition, but this applied only to the Level 5 assessment.

Research question 2: *Is narrative length affected by experimental condition?*

For means and standard deviations regarding narrative length for both the final and intermittent summaries across conditions and levels of assessment, see Table 34.

Table 34

Average (SD) tokens used in the final and intermittent summaries by participants in the Control and Gesture conditions

	Level	Control	Gesture
Final summary	Level 3	66.82 (31.45)	80.20 (27.44)
	Level 4	84.80 (20.88)	105.43 (25.09)
	Level 5	66.82 (20.67)	93.13 (23.42)
Intermittent summary	Level 3	73.27 (16.65)	101.40 (29.72)
	Level 4	89.00 (33.68)	108.80 (24.78)
	Level 5	99.91 (36.94)	140.50 (34.80)

A 2x3 mixed factorial ANOVA was used to assess the influence of condition (a between-subjects factor) and the three different levels of assessment (a within-subjects factor) on both final summary and intermittent summary lengths.

Final summaries

With respect to the number of tokens used in the final summary, the main effect of condition was not significant, F (1, 14) = 2.883, p = n.s., η_p^2 = 0.171. The main effect of Level of assessment was significant, F (2, 28) = 6.145, p < .01, η_p^2 = 0.305. Pairwise Bonferroni-correct post-hoc t-tests indicated that participants used more tokens in their final summaries when completing the Level 4 assessment than when completing the Level 3 (p < .01) assessment. The interaction between Condition and Level was not significant, F (2, 28) = 0.717, p = n.s., η_p^2 = 0.049 (see Figure 8).

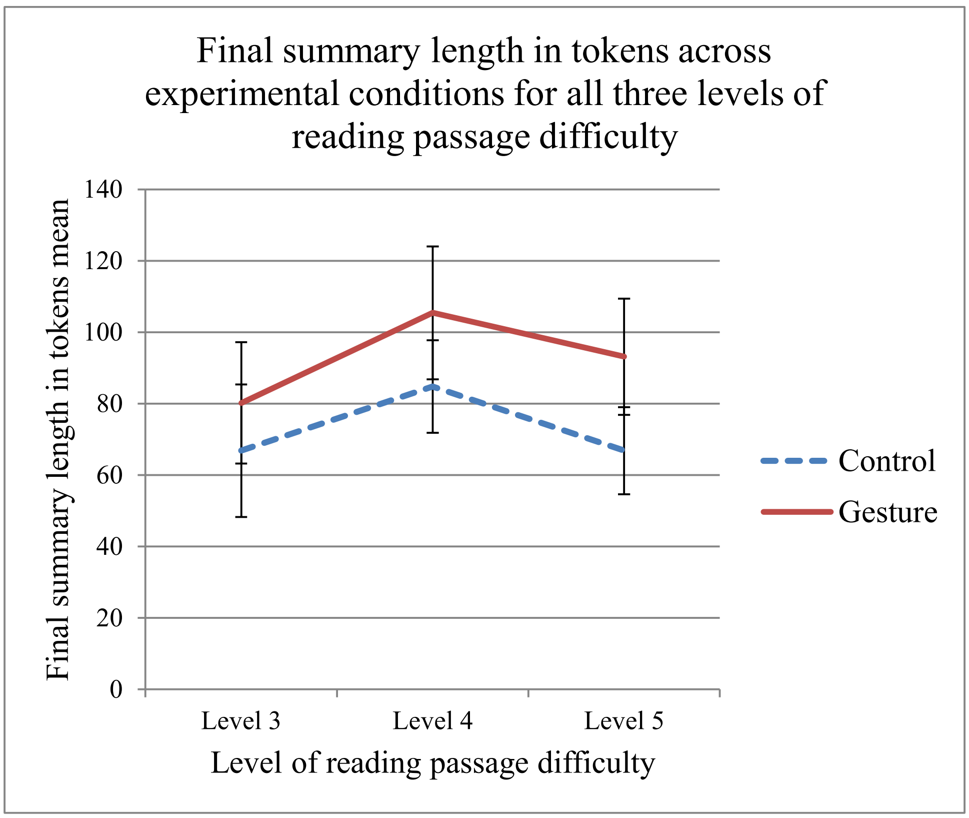

Figure 8. Word tokens used in final summaries compared across conditions and levels of assessment. Error bars represent 95% confidence interval.

Intermittent summaries

With respect to the number of tokens used in the intermittent summaries, the main effect of condition was significant, F (1, 19) = 8.186, $p < .05$, $\eta_p^2 = 0.301$. Participants in the Gesture condition (M = 116.90, SE = 7.46) used significantly more word tokens in their intermittent summaries than participants in the Control condition (M = 87.39, SE = 7.12). The main effect of Level of assessment was significant, F (2, 38) = 10.855, $p < .001$, $\eta_p^2 = 0.364$. Pairwise Bonferroni-correct post-hoc t-tests indicated that participants used more tokens in their final summaries when completing the Level 5 assessment in comparison to when they were completing the Level 3 (p < .01) or Level 4 (p < .05) assessments. The interaction between Condition and Level was not significant, F (2, 38) = 1.069, p = n.s., $\eta_p^2 = 0.053$ (see Figure 9).

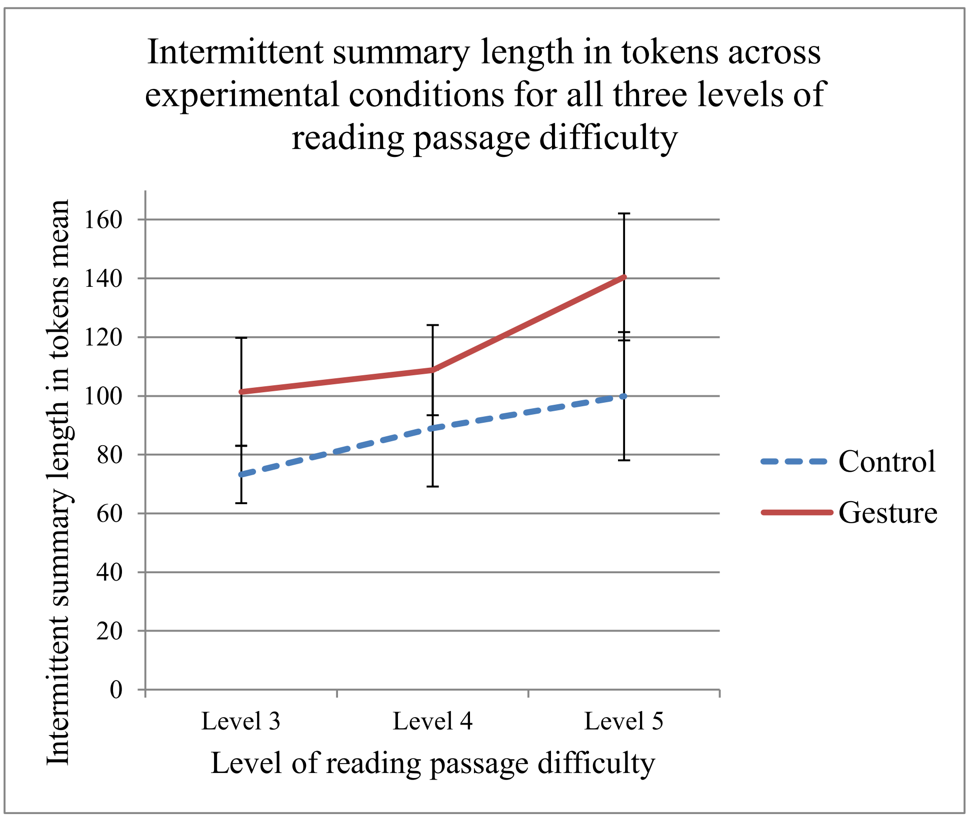

Figure 9. Word tokens used in intermittent summaries compared across conditions and levels of assessment. Error bars represent 95% confidence interval.

Summary of results relevant to research question 2: Is narrative length influenced by experimental condition? The results indicate that when considering the final summaries, experimental condition had no effect. However, when considering intermittent summaries, participants in the Gesture condition used significantly more words than participants in the Control condition.

Research question 3: *Does statistically controlling for intermittent summary length influence the strength of any differences between the Control and Gesture groups on measures of Low Inference and High Inference recall on a standardized reading comprehension assessment?*

In order to address research question 3, multiple regression was performed in order to determine whether performance on Low Inference and High Inference assessments differed significantly across groups and if so, to determine whether the inclusion of intermittent summary length would influence this difference. Since experimental condition is categorical variable and since sample sizes were not equal across groups, dummy coding was used for these analyses. Participants in the Control condition were coded as 0 and participants in the Gesture condition were coded as 1.

In step 1, condition was entered into the regression using forward entry. In step 2, intermittent summary length was entered into the regression. For level 3 and 4 High and Low assessments, the dummy variable was not a significant predictor and therefore will not be reported. Differences according to condition were only seen on the level 5 assessment.

Low Inference scores on the NARA assessment

In Step 1, the entry of the dummy variable led to a statistically significant model $F\ (1, 19) = 6.244$, $p < .05$ and explained 20.8% of the variance in Low Inference scores on the NARA assessment (see Table 35). Participants in the Gesture condition had significantly higher scores than participants in the Control condition. After the entry of intermittent summary tokens, the model explained 22.1% of the variance in Low Inference scores [$F\ (2, 18) = 3.838$, $p < .05$]. The addition of intermittent summary tokens did not explain significantly more variance in Low Inference scores when controlling for the influence of condition [$\Delta R^2 = .052$; $F\ (1, 18) = 1.325$, p = n.s.]. In the final model, neither condition ($\beta = .363$, p = n.s.) nor intermittent summary tokens ($\beta = .264$, p = n.s.) were statistically significant.

Table 35

Regression model predicting Low Inference performance on the NARA assessment

	R	R^2	$Adj.\ R^2$	ΔR^2	B	SE	β	t
Step 1	.497	.247	.208					
Condition					28.295	11.323	.497	2.499*
Step 2	.547	.299	.221	.052				
Condition					20.632	13.053	.363	1.581
Tokens (I)					0.189	0.164	.264	1.151

*$p < .05$; **$p < .01$; ***$p < .001$

High Inference scores on the NARA assessment

In Step 1, the entry of the dummy variable led to a statistically significant model F (1, 19) = 5.457, $p < .05$ and explained 18.2% of the variance in High Inference scores on the NARA assessment (see Table 36). Participants in the Gesture condition had significantly higher scores than participants in the Control condition. After the entry of intermittent summary tokens, the model explained 14.6% of the variance in High Inference scores on the NARA assessment [F (2, 18) = 2.709, p = n.s.]. The addition of intermittent summary tokens did not explain significantly more variance in High Inference scores when controlling for the influence of condition [ΔR^2 = .008; F (1, 18) = 0.193, p = n.s.]. In the final model, condition (β = .526, $p < .05$) was a significant predictor of High Inference scores on the NARA assessment, but intermittent summary tokens (β = -.105, p = n.s.) was not.

Table 36

Regression model predicting High Inference performance on the NARA assessment

	R	R^2	Adj. R^2	ΔR^2	B	SE	β	t
Step 1	.472	.223	.182					
Condition					29.318	12.550	.472	2.336*
Step 2	.481	.231	.146	.008				
Condition					32.657	14.911	.526	2.190*
Tokens (I)					-0.082	0.187	-.105	-0.439

*$p < .05$; **$p < .01$; ***$p < .001$

Summary of results relevant to research question 3: Does statistically controlling for intermittent summary length influence the strength of any differences between the Control and Gesture groups on the NARA assessment? The results show that participants in the Gesture condition have higher Low Inference and High Inference scores on the NARA assessment in comparison to participants in the Control condition. When statistically controlling for the use of intermittent word tokens used, the condition differences persist only when considering High Inference outcomes.

Study 2 Discussion

In Study 1, it was proposed that encouraging gesture production may only be useful for reading comprehension outcomes when individuals struggle to maintain and manipulate information in verbal working memory. In the current study the usefulness of gesture production among children was assessed. Across development the functional capacity of verbal working memory increases (Gathercole et al., 2004). It was predicted then that among children the verbal working memory strain required for this task would be higher in comparison to adults. Additionally, reading passages of three levels of difficulty were used in the current study to further explore the possibility that encouraging gesture production facilitates reading comprehension outcomes when verbal working memory resources are strained. Three research questions were investigated in this study:

Research question 1: *Does gesture production influence reading comprehension outcomes on Low Inference and High Inference assessments?*

It was predicted that participants in the Gesture condition would show advantages over participants in the Control condition on reading comprehension outcomes and that these advantages would be most apparent when assessing performance on the most difficult assessment (Level 5). On the NARA assessment, participants in the Gesture condition did perform significantly better than participants in the Control condition lending support to the hypothesis. However, in contrast to the hypothesis, the gestural advantage was not most salient on the Level 5 assessment. Participants in the Gesture condition had better reading comprehension outcomes than participants in the Control condition regardless of whether they were answering Low or High Inference questions and regardless of which level of assessment they were completing.

When considering the ratings of participant narratives, participants in the Gesture condition again demonstrated an advantage on reading comprehension outcomes: they had higher Main idea construction and Informative ratings than participants in the Control condition and they had lower Recall error ratings than participants in the control condition. Two condition differences were only seen on the Level 5 assessment. Participants in the Gesture condition had higher Understanding scores and lower Added detail scores than participants in the Control condition only when they completed the level 5 assessment.

These results demonstrate that among children, encouraging the use of gesture production can enhance performance on reading comprehension assessments.

Research question 2: *Is narrative length affected by experimental condition?*

In Study 1, adults who gestured spontaneously relayed significantly longer narratives than individuals who did not gesture spontaneously. It was predicted that children in the Gesture condition would relay longer narratives than children in the Control condition and that this would be especially pronounced when examining narrative length on the Level 5 assessment. Participants in the Control condition told stories of comparable length to participants in the Gesture condition when considering final summaries. When considering intermittent summaries, participants in the Gesture condition used significantly more word tokens than participants in the control condition. Neither of the interactions between condition and level were significant, indicating that the hypothesis that any difference across conditions would be most pronounced in Level 5, was not supported.

It is unclear why participants in the Gesture condition relayed longer intermittent summaries than participants in the Control condition, when the final summaries were of comparable length. One possibility is that gesture production only results in an increase in narrative length when participants describe novel information. When children relayed intermittent summaries, they were talking about information that they had just read. When children relayed final summaries, they were talking about the information that they had previously summarized. Future research is required in order to investigate this possibility.

Research question 3: *Does statistically controlling for intermittent summary length influence the strength of any differences between the Control and Gesture groups on measures of Low Inference and High Inference recall on a standardized reading comprehension assessment?*

In Study 1, every advantage among individuals who gestured spontaneously disappeared with statistically controlling for intermittent summary length. It was predicted that similar results would emerge among the children.

The results indicate that statistically controlling for intermittent narrative length only weakened the difference across conditions on the Low Inference scores. When controlling for intermittent narrative length, the difference across groups remained robust when considering High Inference scores. This finding suggests gesture production may be contributing to performance on Low Inference reading comprehension outcomes by supporting the generation of longer narratives. By producing longer intermittent summaries, participants may be more likely to repeat information that they read in the passage and this may facilitate performance on Low Inference reading comprehension assessments. Gesture production seems to be contributing to High Inference reading comprehension outcomes in a different way that is not contingent upon intermittent summary length. It may be the case that gestures serve as a useful way to more robustly encode information in memory. Recall that in the study by Stevanoni and Salmon (2005), the authors proposed that gesture production may enhance memory by serving as props that reduce processing demands and facilitate recall, or by enhancing motivation. In the current study it is possible that both of these factors enhanced memory recall among children in the Gesture condition and that these factors contributed to High Inference reading comprehension outcomes.

Study 2 Conclusion

In conclusion, the results from the current study suggest that encouraging gesture production among children may be beneficial for both High Inference and Low Inference reading comprehension outcomes. Though gesture production seems to be beneficial for reading comprehension outcomes partly because gestures lead to the generation of longer narrative summaries, gestures may also be beneficial by facilitating the imagistic encoding of information in memory. One limitation of the current study is statistical power. The sample sizes used in the analyses were small and as a consequence, the interpretations of the results presented here can only be made tentatively. Future research is required in order to determine whether the pattern of results identified in this study can be replicated using larger samples.

CHAPTER IV

General Discussion

The goal of reading comprehension is to gain an overall understanding of the information described in text (Woolley, 2011). When understanding a text, individuals are thought to rely on mental models of the text ideas. Kintsch (1998) argues that a text-based model (consisting of propositions contained within the text) and a situation model (consisting of inferences about the meaning of the text) are two approaches to understanding the information presented in text. Woolley (2011) has argued that the primary difference between the text-based and situation models is that the situation model requires readers to generate more inferences than the text-based model. Sadoski et al. (1991) have argued that the text-based and situation models described by Kintsch (1998) are compatible with the Dual Coding theory of reading comprehension. The Dual Coding theory of reading comprehension asserts that information from the text can be represented in a verbal code or an imagistic code. Evidence in favor of these theories has been generated. In research by Johnson-Glenberg (2000), individuals were trained to use verbal strategies or visuospatial strategies in a reading comprehension task. Individuals who were trained to rely on verbal strategies excelled on reading comprehension assessments that required the recall of information that was explicitly stated in text (i.e., information that required the generation of few inferences). This suggests that training that supports the verbal code may also support a text-based model of reading comprehension. Individuals who were trained to rely on visuospatial strategies excelled on reading comprehension assessments that required individuals to make inferences to arrive at a correct answer. This suggests that training that supports the visuospatial code may also support a situation model of reading comprehension. Previous research suggests that gesture production is strongly associated with information processing in a visuospatial code (e.g., Morsella et al., 2004; Rauscher et al., 1996) and that the restriction of gesture production may reduce processing in a visuospatial code (e.g. Rimé et al., 1984). In the current study, it was of interest to investigate whether gesture production supports the imagistic code of reading comprehension. It was reasoned that reliance upon a visuospatial code of reading comprehension would be revealed if gesture production was associated with advantages on reading comprehension assessments requiring the generation of inferences. It was also reasoned that reliance upon a verbal code of reading comprehension would be revealed if gesture restriction was associated with advantages on reading comprehension assessments not requiring the generation of inferences. Since verbal and visuospatial strategy preference differs across development (Palmer, 2000), it was of interest to explore whether these predictions would be supported among both adults (Study 1) and children (Study 2). In order to test these predictions, both low and high inference measures were developed. The Low Inference measures were designed to capture participants' ability to recall information that was stated explicitly in the text. For example, the fill-in-the-blank assessments required participants to retrieve specific words that were used in the reading passages. The High Inference measures were designed to capture participants' ability to generate inferences about what was read in the text. The High Inference measures in Study 1 and 2 included the ratings of 7 dimensions of the final summaries that participants generated. Summary generation requires individuals to identify the main ideas presented in a text, to evaluate the importance of the information presented in a text, and to create a coherent summary of what was read (Pecjak, Podlesek, & Pirc, 2011). The use of narrative ratings as an index of High Inference reading comprehension was contingent upon the assumption that the generation of inferences was required for effective summary generation.

The purpose of the current research was to gain a better understanding of how gesture use during summary generation, influences reading comprehension. In the following paragraphs, gesture production will be discussed in terms of its possible influence upon cognitive strategies during a reading comprehension task, and in terms of its possible usefulness in a reading comprehension task.

Gesture production and cognitive strategies

Reading comprehension outcomes differ depending upon whether a verbal or visuospatial strategy is adopted (Johnson-Glenberg, 2000). The purpose of Study 1 was to investigate whether gesture production influences the use of verbal and visuospatial strategies in the context of a reading comprehension task. This is an important area of research since a variety of methodologies are currently used to assess reading comprehension among both children and adults and none of these methods of assessment consider the influence of gesture production. Previous research addressing strategy use during a problem-solving task has demonstrated that gesture production is associated with the adoption of a visuospatial strategy while gesture restriction is associated with the adoption of a verbal strategy (Alibali et al., 2011). The results from Study 1 revealed that among participants who were restricted from gesturing, vocabulary was the best predictor of reading comprehension outcomes, suggesting that these participants may have adopted a verbal strategy. Both visuospatial working memory and motivation scores were the strongest predictors of reading comprehension outcomes among participants who were free to gesture, while motivation was the strongest predictor of reading comprehension outcomes among participants who were encouraged to gesture. Rather than adopting a strictly visuospatial code, participants who were free to or encouraged to gesture seem to have relied upon motivational factors to complete the task.

The finding that movement influences strategy adoption in the context of a reading comprehension task has significant theoretical and practical implications. There is conflicting evidence concerning whether or not measures of verbal and visuospatial working memory predict reading comprehension outcomes. Some research shows that verbal working memory measures predict reading comprehension outcomes (e.g., Cain, Oakhill, & Bryant, 2004; Swanson & Howell, 2001), while other research does not (e.g., Dufva, Niemi, & Voeten, 2001; Oakhill et al., 1986). Similarly, some research shows that visuospatial working memory measures predict reading comprehension outcomes (e.g., Goff et al., 2005; St. Clair-Thompson, 2007) whereas other research does not (e.g., Seigneuric et al., 2000; Swanson & Howell, 2001). The results from the current study suggest that procedural differences in research methodologies may be a key to understanding this lack of convergence. This finding is of practical significance for researchers who investigate the use of oral strategies as a technique to improve reading comprehension (Chinn, Anderson, & Waggoner, 2001; Dennis & Moldof, 1983; McKeown et al., 2009; Nystrand, 1997; Saunders & Goldenberg, 1999). Gesture production is likely to occur among readers who speak about a reading passage and this may lead individuals to adopt a visuospatial or motivational strategy. This finding is also of practical significance for neurological researchers who investigate reading comprehension. A great deal of research concerning reading comprehension is conducted in functional magnetic resonance imaging (fMRI) scanners where individuals are asked to remain still as they complete reading comprehension tasks (e.g., Berl et al., 2010; Humphreys & Gennari, 2014; Landi et al., 2013; Moss et al., 2011). The results of the current study suggest that this methodology may lead individuals to adopt a verbal strategy rather than a visuospatial or motivational strategy thereby biasing research in this field.

Is gesture production useful in a reading comprehension task?

The results from Study 1 revealed that participants in the Gesture condition did not show any notable advantages over participants in the other experimental conditions with respect to reading comprehension performance. However, gesture production is associated with different reading comprehension outcomes when considering individuals in the Control condition who spontaneously gestured. The spontaneous gesturers had significantly higher Fill-in-the-blank (standard) scores in comparison to spontaneous non-gesturers. Additionally, spontaneous gesturers had higher Main idea construction, Summary integration, Understanding, and Informative ratings. Furthermore, spontaneous gesturers had lower Recall error ratings in comparison to spontaneous non-gesturers. Why might spontaneous gesturers show an advantage on reading comprehension outcomes when encouraging gesture production does not have this effect? One possible explanation relates to the different working memory profiles of the spontaneous gesturers and non-gesturers. Iconic gesture production was negatively associated with the measure of verbal working memory. Therefore, individuals who had weaker abilities to store and manipulate verbal information were most likely to gesture. As a result, these participants may have relied upon gesture production as a compensatory mechanism to facilitate lexical retrieval (e.g., Frick-Horbury & Guttentag, 1998) and the verbalization of ideas (e.g., Alibali et al., 2000). The fact that spontaneous gesturers told significantly longer narratives than spontaneous non-gesturers lends support to this possibility.

It was reasoned that if encouraging gesture production in a reading comprehension task is only useful when verbal working memory resources are strained, then encouraging gesture production among children may be associated with reading comprehension advantages (since children have lower verbal working memory abilities in comparison to adults). In line with this prediction, results from Study 2 revealed that encouraging gesture use among children resulted in an advantage in performance on Low Inference and High Inference measures of reading comprehension in comparison to individuals who were not encouraged to gesture.

A notable finding from the analyses is that all spontaneous gesturer advantages on Low Inference and High Inference measures of reading comprehension disappeared when controlling for narrative length among the adults. However, among children the gestural advantage on High Inference measures remains intact when controlling for intermittent summary tokens. This finding suggests that telling longer narrative summaries is not the factor responsible for enhanced reading comprehension outcomes on High Inference measures among children. Future research is required in order to investigate how gesture production might be beneficial for High Inference reading comprehension outcomes among children.

Limitations and Future Directions

There are some limitations regarding the research conducted in Studies 1 and 2. With respect to Study 1, the analyses that compare reading comprehension outcomes among the gesturers and the non-gesturers in the Control condition suffer from low power. Similarly, all of the analyses conducted in Study 2 have limited power. This suggests that the probability of failing to reject the null hypothesis is quite high in both cases. This may have an effect upon the interpretation of many of the research findings. For example, it may be the case that because of the limited sample size in Study 2, the importance of the interaction between the level of reading comprehension passage difficulty and experimental condition is being underestimated. The effect sizes suggest that this might be the case. Though power was quite low, a consistent pattern of results emerged both among the gesturers and non-gesturers in Study 1, and the children in Study 2. This consistency in results coupled with the strong effect sizes suggests that gesture production may play an important role in reading comprehension under particular conditions. In order to gain an understanding of how gesture use during summarization may influence reading comprehension, future research using larger sample sizes is required.

Another limitation of the current study is that the texts used both in Studies 1 and 2 referred to content that was concrete and therefore fairly easy to imagine. It is not clear whether the patterns of results found the in the current study would generalize to reading passages that were more abstract in terms of content. Future research is recommended in order to investigate this.

Finally, the generation of inferences in the current study was conceptualized in terms of degree (using low and high inference measures). It is important to note that a variety of different methods have been used in previous research in order to study the generation of inferences during reading comprehension. These methods have relied upon studying the different *types* of inferencing (i.e., referential and thematic inferences) (see Graesser, Singer, & Trabasso, 1994, for a review). Future research is required in order to investigate how reading comprehension is differentially influenced by the degree and kind of inferences generated by the reader.

Conclusion

In conclusion, gestures that are encouraged or produced spontaneously may enhance reading comprehension outcomes among individuals whose verbal resources are strained. Future research is required in order to investigate whether encouraging adults to gesture while recalling details from difficult reading passages would result in an advantage on reading comprehension measures in comparison to adults who are provided with no instructions regarding their hand movements. Future research is also required in order to investigate whether the beneficial effect of gesture production can be generalized to children who are first learning how to read.

The findings generated by this research may be useful to parents and teachers of children who are developing their reading skills. When children struggle with difficult words and difficult concepts in reading passages, both parents and teachers could encourage children to use their hands to represent ideas in meaningful ways. This provides a useful strategy for children to compensate for difficulties representing information verbally and may lead to improved performance on both Low Inference and High Inference assessments of reading comprehension.

References

Alibali, M. W. (2005). Gesture in visuospatial cognition: Expressing, communicating, and thinking about visuospatial information. *Visuospatial Cognition & Computation, 5*, 307-331.

Alibali, M.W., Kita, S., Young, A.J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes, 15 (*6), 593-613.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science, 22* (9) 1138–1144.

Alloway, T.P. (2007). *Automated Working Memory Assessment*. London, UK: Harcourt Assessment.

Alloway, T.P. Gathercole, C.W., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87, 85-106.*

Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences, 19*, 252-256.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Hove: Erlbaum.

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

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-609.

Bashir, A.S., & Hook, P.E. (2009). Fluency: A key link between word identification and comprehension. *Language, Speech & Hearing Services in Schools, 40* (2), 196-200.

Begg, I., & Anderson, M.C. (1972). Imagery and associative memory in children. *Journal of Experimental Child Psychology, 21* (3), 480-489.

Berl, M.M. et al. (2010). Functional anatomy of listening and reading comprehension during development. *Brain & Language, 114*, 115-125.

Bishop, D.V.M. (1997). *Uncommon understanding: Development and disorder of language comprehension in children.* Hove: Psychological Press.

Block, C.C. (2004). *Teaching comprehension: The comprehension process approach.* Boston: Pearson.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*, 44–48.

Bowers, P.G., & Sunseth, K. (2002). Rapid naming and phonemic awareness: Contributions to reading, spelling, and orthographic knowledge. *Scientific Studies of Reading, 6*, 401-429.

Broaders, S.C., Wagner-Cook, S., Mitchell, Z., Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136* (4), 539-550.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory, 14* (5), 553-569.

Cain, K. (2007). Deriving word meanings from context: Does explanation facilitate contextual analysis? *Journal of Research in Reading, 30* (4), 347-359.

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal, 13*, 31-56.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* (1), 31-42.

Cattaneo, Z., Fastame, M.C., Vecchi, T., & Cornoldi, C. (2006). Working memory, imagery and visuospatial mechanisms. In T. Vecchi, & G. Bottini (Eds.) *Imagery and Visuospatial Cognition*. The Netherlands: John Benjamins Publishing Company.

Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visuospatial imagery training on the reading and listening comprehension of low listening comprehenders in year 2. *Journal of Research in Reading, 22* (3), 241-256.

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

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*, 580-601.

Chinn, C.A., Anderson, R.C., & Waggonner, M.A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly, 36* (4), 378-411.

Crammond, D. J. (1997) Motor imagery: Never in your wildest dreams. *Trends in Neuroscience, 20*, 54–57

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

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1–18.

Daneman, M. & Merikle, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3* (4), 422-433.

Decety, J., Jeannerod, M., Durozard, D., & Baverel, G. (1993) Central activation of autonomic effectors during mental simulation of motor actions in man. *Journal of Physiology,461*, 549-563.

Dennis, R. P. & Moldof, E.P. (1983). *A handbook on interpretive reading and discussion*. Chicago: Great Books Foundation.

de Ruiter, J.-P. (1998). *Gesture and speech production.* Unpublished doctoral dissertation, Katholieke Universiteit Nijmegen, Nijmegen, The Netherlands.

Dufva, M., Niemi, P., & Voeten, M.J.M. (2001). The role of phonological memory, word recognition, and comprehension skills in reading development: from preschool to grade 2. *Reading and Writing: An Interdisciplinary Journal, 14,* 91-117.

Dunn, L.M., & Dunn, L.M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test-Third Edition.* Circle Pines, MN: American Guidance Service.

Farah, M.J. (1995). Current issues in the neuropsychology of image generation. *Neuropsychologia, 33,* 1455-1471.

Feyereisen, P. & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior, 23* (2), 153-171.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *American Journal of Psychology, 111*, 43-62.

Gainotti, G., Silveri, M. C., Daniele, A. & Giustolisi, L. (1995) Neuroanatomical correlates of category-specific semantic disorders: A critical survey. *Memory, 3*, 247–64.

Gathercole, S.E., Pickering, S.J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40* (2), 177-190.

Gersten, R., Fuchs, L.S., Williams, J.P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research, 71,* 279-320.

Glenberg, A.M., Brown, M., & Levin, J.R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology, 32, 389-399.*

Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., & Kaschak, M.P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96* (3), 424-436.

Glenberg, A.M., & Kaschak, M.P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9* (3), 558-565.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes, 28*, 1–26.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language, 43*, 379-401.

Goff, D.A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing, 18,* 583-616.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101* (3), 371-395.

Guillot, A., & Collet, C. (2005). Contribution from neurophysiological and psychological methods to the study of motor imagery. *Brain Research Reviews, 50*, 387-397.

Harm, M.W., & Seidenberg, M.S. (1999). Phonology, reading, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491- 528.

Hostetter, A.B., & Alibalil, M.W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture, 7* (1), 73-95.

Hostetter, A.B., Alibali, M.W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15* (3), 495-514.

Hostetter, A.B., & Skirving, C.J. (2011). The effect of visual vs. verbal stimuli on gesture production. *Journal of Nonverbal Behavior, 35* (3), 205-223.

Humphreys, G.F. & Gennari, S.P. (2014). Competitive mechanisms in sentence processing: Common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage, 84*, 354-366.

Iverson, J.M., & Braddock, B.A. (2011). Gesture and motor skill in relation to language in children with language impairment. *Journal of Speech, Language, and Hearing Research, 54*, 72-86.

Jeannerod, M. (1994) The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*, 187–245.

Jeannerod, M. (1995) Mental imagery in the motor context. *Neuropsychologia, 33*, 1419–32.

Jincho, N., Namiki, H., & Mazuka, R. (2008). Effects of verbal working memory and cumulative linguistic knowledge on reading comprehension. *Japanese Psychological Research, 50* (1), 12-23.

Johnson-Glenberg, M.C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visuospatial strategies. *Journal of Educational Psychology, 92 (4), 772-782.*

Juel, C. (1998). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 417-447.

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99* (1), 122-149.

Kelly, S., Byrne, K., & Holler, J. (2011). Raising the ante of communication: Evidence for enhanced gesture use in high stakes situations. *Information, 2*, 579-593.

Kendeou, P., Savage, R., & Van den Broek, P. (2009). Revising the simple view of reading. *British Journal of Educational Psychology, 79,* 353-370.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist, 49,* 294-303.

Kirby, J. R. (1993). Collaborative and competitive effects of verbal and visuospatial processes. *Learning and Instruction, 3,* 201-214.

Klauda, S.L., & Guthrie, J.T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100* (2), 310-321.

Kolic-Vehovec, S., Bajsanski, I., & Zubkovic, B.R. (2011). The role of reading strategies in scientific text comprehension and academic achievement of university students. *Review of Psychology, 18* (2), 81-90.

Kosslyn, S.M. (1988). Imagery in learning. In *Perspectives in memory research,* ed. M.S. Gazzaniga, 245-273. London: MIT Press.

Kosslyn, S.M., M. Behrmann, & Jeannerod, M. (1995a). The cognitive neuroscience of mental imagery. *Neuropsycholigia, 33*, 1335-1344.

Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. (1995) Topographical representations of mental images in primary visuospatial cortex. *Nature, 378*, 496– 98.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). New York: Cambridge University Press.

LaBerg, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Landi, N., Frost, S.J., Mencl, W.E., Sandak, R., & Pugh, K.R. (2013). Neurological bases of reading comprehension: Insights from neuroimaging studies of word-level and text-level processing in skilled and impaired readers. *Reading & Writing Quarterly, 29*, 145-167.

Langer, J.A. (1990). Understanding literature. *Language Arts, 67*, 812-816.

Langer, J.A. (1995). *Envisioning Literature: Literacy Understanding and Literature Instruction.* New York: Teachers College Press.

Lau, K., & Chan, D.W. (2003) Reading strategy use and motivation among Chinese good and poor readers in Hong Kong. *Journal of Research in Reading, 26* (2), 177-190.

Logie, R.H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology, 38A*, 229-247.

Long, S.A., Winograd, P.N., & Bridge, C.A. (1989). The effects of reader and text characteristics on imagery reported during and after reading. *Reading Research Quarterly, 24,* 353-372.

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates

Manheim, R (Ed.) (1977). *Grimms' tales for young and old.* Garden City, NY: Doubleday.

Mar, R.A. (2003). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia, 42*, 1414-1434.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86,* 389-401.

McKeown, M.G., Beck, I.L., & Blake, R.G.K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44* (3) 218-253.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Morrel-Samuels, P., & Krauss, R.M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 615-623.

Morsella, E., & Krauss, R.M. (2004). The role of gestures in visuospatial working memory and speech. *The American Journal of Psychology, 117,* 411-424.

Moss, J., Schunn, C.D., Schneider, W., McNamara, D., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: *Cognitive control or discourse comprehension. NeuroImage, 58*, 675-686.

Nation, K. Adams, J.W., Bowyer-Crane, C.A., & Snowling, M.J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73,* 139-158.

Neal, J.C., & Kelly, P.R. (2002). Delivering the promise of academic success through late intervention. *Reading and Writing Quarterly, 18,* 101-117.

Neale, M.D. (1999). *Neale Analysis of Reading Ability*: Manual (3rd ed.). Australian Council for Educational Research: Melbourne, Australia.

Nichols, W.D., Rupley, M.H., & Rasinnski, T. (2009). Fluency in learning to read for meaning: Going beyond repeated readings. *Literacy Research and Instruction, 48,* 1-13.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.

Oakhill, J., & Cain, K. (2007). Introduction to comprehension development. In *Children's comprehension problems in oral and written language: A cognitive perspective,* ed. K. Cain and Oakhill, J., 3-40. London: The Guilford Press.

Oakhill, J.V., Cain, K., & Bryant, P.E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18* (4), 443-468.

Oakhill, J., Yuill, N., & Parkin, A. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading, 9* (2), 80-91.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Holt, Rinehart & Winston.

Palmer, S. (2000). Development of phonological recoding and literacy acquisition: A four-year cross-sequential study. *British Journal of Developmental Psychology, 18*, 533-555.

Parrill, F., Bullen, J., & Hoburg, H. (2010). Effects of input modality on speech-gesture integration. *Journal of Pragmatics, 42*, 3130-3137.

Pearson, P.D., Hiebert, E.H., & Kamil, M.L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42* (2), 282-296.

Pearson, P.D. & Johnson, D.D. (1978). *Teaching reading comprehension.* New York: Holt, Rinehart and Winston.

Pecjak, S., Podlesek, A., & Pirc, T. (2011). Model of reading comprehension for 5th grade students. *Studia Psychologica, 53* (1), 53-67.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego: Academic Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11* (4), 357-383.

Pine, K.J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10* (6), 747-754.

Plass, J. L., Chun, D. D., Mayer, R. E., & Leutner, D. (1998). Supporting visuospatial and verbal learning preferences in second-language multimedia learning environment. *Journal of Educational Psychology, 90,* 25-36.

Pulvermüller, F. (1999) Words in the brain's language. *Behavioral and Brain Sciences, 22* (2), 253–336.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-231.

Rimé, B., Shiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion, 8*, 311–325.

Sadoski, M. (1983). An exploratory study of the relationship between reported imagery. *Reading Research Quarterly, 19,* 110-123.

Sadoski, M. (1999). Comprehending comprehension. *Reading Research Quarterly, 34* (4), 493-500.

Sadoski, M., Paivio, A., Goetz, E.T. (1991). A critique of schema theory in reading and a dual coding alternative. *Reading Research Quarterly, 26*, 463-484.

Sadoski, M., & Quast, Z. (1990). Reader response and long-term recall for journalistic text: The roles of imagery, affect, and importance. *Reading Research Quarterly, 25,* 256-272.

Saunders, W.M., & Goldenberg, C. (1999). Effects of instructional conversations and literature logs on limited- and fluent-English-proficient students' story comprehension and thematic understanding. *The Elementary School Journal, 99* (4), 277-301.

Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior, 27,* 1-14.

Seigneuric, A., Ehrlich, M.-F., Oakhill, J.V., & Yuill, N.M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 13,* 81-103.

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

Smithson, L., & Nicoladis, E. (2013). Verbal memory resources predict iconic gesture use among monolinguals and bilinguals. *Bilingualism: Language and Cognition, 16* (4), 934-944.

Smithson, L., & Nicoladis, E. (2014). Lending a Hand to Imagery? The Impact of Visuospatial Working Memory Interference Upon Iconic Gesture Production in a Narrative Task. *Journal of Nonverbal Behavior*. Advance online publication. doi: 10.1007/s10919-014-0176-2.

Smithson, L., & Nicoladis, E. (under review). Visuospatial working memory capacity predicts physiological arousal in a narrative task.

St. Clair-Thompson, H.L. (2007). The influence of strategies upon relationships between working memory and cognitive skills. *Memory, 15* (4), 353-365.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-407.

Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior, 29* (4), 217-233.

Stull, A., & Mayer, R.E. (2007). Learning by doing versus learning by viewing: Three experimental comparisons of learner-generated versus author-provided graphic organizers. *Journal of Educational Psychology, 99* (4), 808-820.

Swanborn, M.S.L., & de Glopper, K. (2002). Impact of reading prose on incidental word learning from context. *Language Learning, 52,* 95-117.

Swanson, H.L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology, 93* (4), 720-734.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Taboada, A., Tonks, S.M., Wigfield, A., & Guthrie, J.T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing, 22*, 85-106.

Thiede, K.W., & Anderson, M.C.M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129-160.

Trbovich, P.L., & LeFevre, J-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31* (5), 738-745.

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *Journal of Educational Research, 100* (2), 81-101.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology, 31,* 142-166.

Wade, S.E., Buxton, W.M., & Kelly, M. (1999). Using think-alouds to examine reader text interest. *Reading Research Quarterly, 34*, 194-216.

Wang, J. H. Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. Reading Research Quarterly, 39, 162–186.

Weaver, C. A., & Bryant, D.S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition, 23* (1), 12-22.

Wesp, R., Hesse, J., Keutmann, D. & Wheaton, K. (2001). Gestures maintain visuospatial imagery. *American Journal of Psychology, 114*, 591–600.

West, R.L., Morris, C.W., & Nichol, G.T. (1985). Chapter 2: Spatial cognition on nonspatial tasks: Finding spatial knowledge when you're not looking for it. In R. Cohen (Ed.) *The Development of Spatial Cognition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Wigfield, A. & Guthrie, J.T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420-432.

Woolley, G. (2011). *Reading comprehension: Assisting children with learning difficulties*. Dordrecht, The Netherlands: Springer International.

Worthy, J., Patterson, E., Salas, R., Prater, S., & Turner, M. (2002). More than just reading: The human factor in research resistant readers. *Reading Research and Instruction, 41, 177-202.*

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E. & Evans, A. C. (1996) Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*, 29–46.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), The psychology of learning and motivation (Vol. 44, pp. 35–62). New York: Academic Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

Appendix A

Practice passage (‘The Mouse, the Bird, and the Sausage’)

<u>Question:</u>

1) The bird’s job was to fly to the ______________ every day and bring wood.

Level 1 (‘Old Sultan’)

<u>Questions:</u>

1) The peasant wanted to shoot Sultan because he didn’t have any ______________ left.

2) While the peasant and his wife were at work, they left their ______________ behind the hedge.

3) The wolf was planning to take the baby when the peasant and his wife were making ______________.

4) The father ______________ when he saw the wolf running across the fields with his baby.

5) The peasant asked his wife to give the dog some ______________ to eat.

6) The wolf asked Sultan to look the other way when he stole a fat ______________ now and then.

7) Sultan told the ______________ about the wolf’s plans.

8) It looked like the cat was picking up ______________.

9) It looked like the cat’s tail was a ______________.

10) The boar revealed the location of the wolf after the cat ______________ his ear.

Level 2 (‘The Owl’)

<u>Questions:</u>

1) A horned ______________ from a nearby forest somehow got into a barn

2) The master said that the hired man would pick up a ______________ before getting too close to a dead chicken.

3) The townspeople came running with pikes, ______________, scythes, and axes.

4) The hero climbed a ______________ to reach the owl.

5) The owl rolled its eyes, puffed up its feathers, and gnashed its ______________.

6) Just before the hero started to back down he started to ______________ and nearly fainted.

7) The townspeople were saying “______________ home”, as they egged the hero on.

8) The townspeople thought that the owl had ______________ and mortally wounded the strongest man in the town.

9) The mayor said that allowance would be made for the ______________, straw, and hay that was stored in the barn.

10) The townspeople set fire to the four ______________ of the barn and the owl died.

Appendix B

For each question below please check the rating applies best:

1) Main idea construction: How accurate were the participants in talking about the main ideas presented in the reading passage?

1 point	2 points	3 points
☐	☐	☐
Incomplete	Approximate	Correct and complete

2) Summary integration: Which option below best describes the integration of the summary?

1 point	2 points	3 points
☐	☐	☐
Poorly integrated	Averagely integrated	Well integrated and cohesive

3) How well do you think the participant understood the reading passage?

1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐
Very poorly						Very well

4) How informative was the narrative?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Not informative					Moderately informative					Very informative

5) How much did you enjoy reading the narrative?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
No enjoyment					Moderate enjoyment					A great deal of enjoyment

6) How many details (not included within the original reading passage) were added to the narrative?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
No additional details were added					A moderate amount of additional details were added					Many additional details were added

7) How many recall errors did participants make in relaying their narrative?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
No recall errors					A moderate amount of recall errors					Many recall errors

Appendix C

Table 1

Correlational analyses between predictors and Low and High Inference measures of reading comprehension in the Control condition

Measure	2	3	4	5	6	7	8	9
1) Main idea construction	.834**	.892**	.886**	.669**	-.204	-.837**	.537**	.843**
2) Summary integration	-	.754**	.795**	.638**	-.018	-.742**	.505**	.707**
3) Understanding	-	-	.921**	.749**	-.384*	-.883**	.588**	.943**
4) Informative	-	-	-	.822**	-.156	-.904**	.478*	.907*
5) Enjoyment	-	-	-	-	-.104	-.693**	.390*	.757**
6) Added details	-	-	-	-	-	.193	-.291	-.302
7) Recall errors	-	-	-	-	-	-	-.660**	-.902**
8) Fill-in-the-blank (standard)	-	-	-	-	-	-	-	.573**
9) Accuracy for proposition inclusion	-	-	-	-	-	-	-	-

Table 2

Correlational analyses between predictors and Low and High Inference measures of reading comprehension in the Gesture condition

Measure	2	3	4	5	6	7	8	9
1) Main idea construction	.648**	.789**	.761**	.457*	.041	-.759**	.421*	.700**
2) Summary integration	-	.658**	.753**	.443*	-.039	-.639**	.320	.536**
3) Understanding	-	-	.831**	.388*	-.277	-.936**	.508**	.734**
4) Informative	-	-	-	.616**	-.062	-.732**	.522**	.784**
5) Enjoyment	-	-	-	-	.219	-.403*	.105	.529**
6) Added details	-	-	-	-	-	.280	-.265	.174
7) Recall errors	-	-	-	-	-	-	-.385*	-.649**
8) Fill-in-the-blank (standard)	-	-	-	-	-	-	-	.538**
9) Accuracy for proposition inclusion	-	-	-	-	-	-	-	-

Table 3

Correlational analyses between predictors and Low and High Inference measures of reading comprehension in the Restricted condition

Measure	2	3	4	5	6	7	8	9
1) Main idea construction	.546**	.844**	.781**	.588**	-.047	-.834**	.391*	.788**
2) Summary integration	-	.777**	.702**	.516**	-.214	-.587**	.538**	.604**
3) Understanding	-	-	.893**	.708**	-.116	-.907**	.511**	.876**
4) Informative	-	-	-	.772**	-.146	-.857**	.511**	.829**
5) Enjoyment	-	-	-	-	.026	-.678**	.480*	.650**
6) Added details	-	-	-	-	-	-.003	-.137	.013
7) Recall errors	-	-	-	-	-	-	-.408*	-.888**
8) Fill-in-the-blank (standard)	-	-	-	-	-	-	-	.468*
9) Accuracy for proposition inclusion	-	-	-	-	-	-	-	-

Appendix D

Motivation according to experimental condition

	Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=29)	ANOVA
MRQ	2.96	2.84	2.75	2.61	F (3, 110) = 1.444
Efficacy	(0.66)	(0.61)	(0.74)	(0.66)	p = .234, η_p^2 = 0.038
MRQ	3.04	2.99	2.88	2.74	F (3, 110) = 1.318
Challenge	(0.58)	(0.56)	(0.72)	(0.60)	p = .272, η_p^2 = 0.035
MRQ	3.26	3.17	3.14	3.24	F (3, 110) = 0.296
Curiosity	(0.50)	(0.58)	(0.65)	(0.44)	p = .828, η_p^2 = 0.008
MRQ	3.09	2.95	3.01	2.90	F (3, 110) = 0.442
Involvement	(0.71)	(0.63)	(0.55)	(0.68)	p = .723, η_p^2 = 0.012

Appendix E

Decoding speed according to experimental condition

	Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=29)	ANOVA
Easy P1	91.82	91.57	92.70	93.72	F (3, 110) = 0.188
	(10.41)	(10.34)	(13.81)	(13.72)	p = .905, η_p^2 = 0.005
Easy P2	73.71	75.43	75.78	77.76	F (3, 110) = 0.717
	(9.27)	(8.95)	(12.52)	(10.90)	p = .544, η_p^2 = 0.019
Easy P3	70.64	71.93	73.33	73.31	F (3, 110) = 0.344
	(9.73)	(9.52)	(14.56)	(12.16)	p = .793, η_p^2 = 0.009
	Control (N=28)	Gesture (N=30)	Restricted (N=27)	Written (N=25)	
Standard	96.75	99.13	100.19	96.68	F (3, 106) = 0.355
P1	(11.85)	(12.85)	(17.88)	(18.17)	p = .786, η_p^2 = 0.010
Standard	95.29	97.30	98.22	96.20	F (3, 106) = 0.204
P2	(12.39)	(14.19)	(16.77)	(16.04)	p = .894, η_p^2 = 0.006
Standard	55.86	55.73	57.15	55.76	F (3, 106) = 0.166
P3	(7.18)	(8.28)	(9.44)	(10.03)	p = .919, η_p^2 = 0.005

Appendix F

The purpose of this Appendix is to provide information concerning the results from Study 1 using an alternative approach to the analyses. In Study 1, numerous reading comprehension outcomes were used in the form of Low and High Inference measures. In this Appendix, a composite measure was created for the Low and High Inference measures (based upon the correlations between these outcomes). This was done in an attempt to reduce the incidence of Type 1 errors, and also in order to reduce redundancies in the analyses. See Table 1 for a correlation matrix that shows the associations between Low and High Inference measures of reading comprehension.

Table 1

Correlation matrix showing the association between High and Low Inference measures of reading comprehension across all conditions (N = 109)

Measure	2	3	4	5	6	7	8	9
1) Main idea construction	.694**	.830**	.811**	.529**	-.101	-.788**	.355**	.716**
2) Summary integration	-	.697**	.736**	.488**	-.095	-.662**	.328**	.529**
3) Understanding	-	-	.888**	.553**	-.281**	-.888**	.403**	.783**
4) Informative	-	-	-	.662**	-.175	-.831**	.365**	.768**
5) Enjoyment	-	-	-	-	.015	-.546**	.230*	.597**
6) Added details	-	-	-	-	-	.131	-.263**	-.040
7) Recall errors	-	-	-	-	-	-	-.343**	-.757**
8) Fill-in-the-blank (standard)	-	-	-	-	-	-	-	.498**
9) Accuracy for proposition inclusion	-	-	-	-	-	-	-	-

*p=0.05, **p=0.01

Correlational analyses revealed that the High Inference measures of reading comprehension were all significantly associated with one another, with the exception of the added details rating. In order to obtain a High Inference composite score, the score on the recall errors measurement was reverse coded. Subsequently scores on each of the measures were converted to percentages and then averaged across the 6 measures (excluding added details).

Correlational analyses also revealed that the Low Inference measures of reading comprehension were significantly associated. However, the accuracy for proposition inclusion scores were more highly correlated with the High Inference measures than with the fill-in-the-blank measure. As a result, the only Low Inference measure of reading comprehension used in the current analyses was the fill-in-the-blank measure. This score was converted into a percentage to facilitate comparisons with the High Inference measures (see Table 2).

Table 2

Means and standard deviations for Low and High Inference measures across conditions

	Control	Gesture	Restricted	Written
Low Inference	65.71 (18.94)	64.33 (15.69)	72.22 (13.40)	58.00 (14.72)
High Inference	66.57 (20.64)	65.87 (17.26)	68.47 (17.29)	75.54 (13.00)

Correlational analyses for regression analyses:

Table 3

Correlational analyses relevant for regression analyses in the Control condition

Measure	2	3	4	5	6
1) Low Inference score	.584**	.090	.210	.451*	.418*
2) High Inference score	-	.236	.454*	.280	.242
3) Verbal WM	-	-	.492**	.281	.246
4) Visuospatial WM	-	-	-	.113	.108
5) Total MRQ	-	-	-	-	.090
6) PPVT	-	-	-	-	-

*p=0.05, **p=0.01

Table 4

Correlational analyses relevant for regression analyses in the Gesture condition

Measure	2	3	4	5	6
1) Low Inference score	.450*	.439*	-.052	.614**	.319†
2) High Inference score	-	.023	-.138	.298	.355†
3) Verbal WM	-	-	.517**	.227	-.090
4) Visuospatial WM	-	-	-	-.084	.018
5) Total MRQ	-	-	-	-	.403*
6) PPVT	-	-	-	-	-

*p=0.05, **p=0.01, †p<0.09

Table 5

Correlational analyses relevant for regression analyses in the Restricted condition

Measure	2	3	4	5	6
1) Low Inference score	.514**	.152	.138	.205	.485*
2) High Inference score	-	.153	.236	.099	.379†
3) Verbal WM	-	-	.193	.184	.051
4) Visuospatial WM	-	-	-	-.144	.237
5) Total MRQ	-	-	-	-	.429*
6) PPVT	-	-	-	-	-

*p=0.05, **p=0.01, †p<0.06

Part 1: Predictors of reading comprehension according to experimental condition

In the original analyses, seven factors were included as predictors of reading comprehension. In order to reduce the number of predictors the motivation subscales were totaled and this was used as a unified measure of motivation in the current analyses. Furthermore, rather than using stepwise regression, each of the predictors was entered into the regression equation. In the current analyses PPVT, total motivation, verbal working memory, and visuospatial working memory measures were entered as predictors of reading comprehension.

Hypothesis 1: *When gestures are encouraged and when they are free to be produced spontaneously, the visuospatial working memory measure will be the best predictor of reading comprehension*

Control condition

Table 6

Information concerning regression analyses for Low and High Inference measures among participants in the Control condition

Measure	*Adj. R^2*	*F test*	*Predictor(s)*
Low Inference score	0.304	F (4, 23) = 3.949, p < .05	**Verbal WM:** β = -.260 t (23) = -1.332, p = .196 **Visuospatial WM:** β = .241 t (23) = 1.307, p = .204 **Total MRQ:** β = .459 t (23) = 2.743, p = .012* **PPVT:** β = .414 t (23) = 2.499, p = .020*
High Inference score	0.164	F (4, 22) = 2.274, p = n.s.	

*p=0.05

Gesture condition

Table 7

Information concerning regression analyses for Low and High Inference measures among participants in the Gesture condition

Measure	*Adj.* R^2	*F test*	*Predictor(s)*
Low Inference score	0.478	F (4, 25) = 7.633, p < .001	**Verbal WM:** β = .524 t (25) = 3.043, p = .005** **Visuospatial WM:** β = -.294 t (25) = -1.783, p = .087† **Total MRQ:** β = .383 t (25) = 2.379, p = .025* **PPVT:** β = .217 t (25) = 1.417, p = .169
High Inference score	0.050	F (4, 25) = 1.378, p = n.s.	

*p=0.05, **p=0.01, †p<0.09

Hypothesis 2: *When gestures are restricted, factors associated with a verbal strategy of encoding (i.e., vocabulary and verbal working memory measures) will be the best predictors of reading comprehension*

Table 8

Information concerning regression analyses for Low and High Inference measures among participants in the Restricted condition

Measure	*Adj.* R^2	*F test*	*Predictor(s)*
Low Inference score	0.145	F (4, 21) = 2.058, p = n.s.	
High Inference score	0.004	F (4, 21) = 1.022, p = n.s.	

Summary of results from Part 1: In the original analyses, motivation and visuospatial working memory measures emerged significant predictors of reading comprehension in the Control condition; motivation measures emerged as significant predictors of reading comprehension in the Gesture condition; and vocabulary emerged as a significant predictor of reading comprehension in the Restricted condition. In the current analyses, motivation and vocabulary emerged as significant predictors in the Control condition, and motivation and verbal working memory emerged as significant predictors in the Gesture condition. No significant predictors emerged in the Restricted condition. Across both methods of analyses motivation was a significant predictor of reading comprehension outcomes in the Control and Gesture conditions. In the current analyses, the importance of verbal measures for reading comprehension outcomes in the Control and Gesture conditions was revealed.

Part 2: Performance on Low Inference and High Inference measures of reading comprehension according to experimental condition

Recall that each experimental group was assigned dummy values (*D1*, *D2, D3*): Control condition (0, 0, 0); Gesture condition (1, 0, 0); Restricted condition (0, 1, 0); and Written condition (0, 0, 1). In the current analyses, D_1 refers to the Gesture condition, D_2 refers to the Restricted condition, and D_3 refers to the Written condition.

In the original analyses, stepwise regression was used. In the current analyses, hierarchical regression was used.

Hypothesis 3: *Individuals who are restricted from moving will perform better on Low Inference assessments*

Reading comprehension outcomes across experimental conditions

Hierarchical multiple regression was performed to investigate whether performance on the Low Inference assessments differed across experimental conditions when controlling for intermittent summary length.

Low Inference

In the first step, intermittent summary length was entered into the regression. This model was statistically significant F (1, 107) = 14.753, $p < 0.001$ and explained 11.3% of the variance in Low Inference scores. In the second step, the dummy variables were entered into the multiple regression: D_1, D_2, and D_3. This model was statistically significant F (4, 104) = 5.622, $p < 0.001$ and explained 14.6% of the variance in Low Inference scores. In this model, none of the dummy variables were statistically significant (see Table 9).

Table 9

Regression model predicting Low Inference performance

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.348	.121	.113					
Tokens					0.030	0.008	.348	3.841***
Step 2	.422	.178	.146	.057†				
Tokens					0.028	0.009	.329	3.323**
D1					-3.466	4.095	-.094	-0.846
D2					6.252	4.146	.165	1.508
D3					-3.204	4.424	-.082	-0.724

*$p < .05$; **$p < .01$; ***$p < .001$; †p<0.08

Hypothesis 4: *Individuals who are encouraged to gesture will perform better on High Inference assessments*

High Inference

In the first step, intermittent summary length was entered into the regression. This model was statistically significant F (1, 107) = 38.234, $p < 0.001$ and explained 25.6% of the variance in High Inference scores. In the second step, the dummy variables were entered into the multiple regression: D_1, D_2, and D_3. This model was statistically significant F (4, 104) = 25.920, $p < 0.001$ and explained 48.0% of the variance in High Inference scores. In this model, only the dummy variable associated with the Written condition was statistically significant (see Table 10).

Table 10

Regression model predicting High Inference performance

	R	R^2	$Adj.\ R^2$	ΔR^2	B	SE	β	t
Step 1	.513	.263	.256					
Tokens					0.047	0.008	.513	6.183***
Step 2	.707	.499	.480	.236***				
Tokens					0.068	0.007	.750	9.699***
D1					-6.127	3.394	-.157	-1.805†
D2					0.895	3.437	.022	0.260
D3					19.497	3.667	.471	5.317***

*$p < .05$; **$p < .01$; ***$p < .001$; †p<0.08

Summary of results from Part 2: The original analyses revealed that participants in the Restricted condition demonstrated an advantage on one of the Low Inference measures of reading comprehension. They also revealed that participants in the Gesture condition demonstrated higher scores only on the Added Details measure. Individuals in the Written condition tended to show the best performance on High Inference measures of reading comprehension. In the current analyses, none of the experimental groups showed an advantage on the Low Inference measures of reading comprehension. Just as was seen in the original analyses, participants in the Written condition showed an advantage on the High Inference measure of reading comprehension.

Part 3: Performance on Low Inference and High Inference measures of reading comprehension according to whether individuals gestured spontaneously or not

Hypothesis 5: *Iconic gesture production is negatively associated with the measure of verbal working memory and positively associated with the measure of visuospatial working memory*

In the original analyses, stepwise regression was used in order to investigate hypothesis #5. In the current analyses, hierarchical multiple regression was performed to investigate whether verbal working memory and visuospatial working memory were significant predictors of iconic gesture rate when controlling for intermittent summary length and total motivation.

In the first step, intermittent summary length and motivation were entered into the regression. This model was not statistically significant F (2, 24) = 1.214, p = n.s. and explained 1.6% of the variance in iconic gesture rate. In the second step, verbal working memory and visuospatial working memory were entered into the multiple regression: This model was statistically significant F (4, 22) = 4.040, $p < 0.05$ and explained 31.9% of the variance in iconic gesture rate (see Table 11).

Table 11

Regression model predicting iconic gesture rate

	R	R^2	$Adj. R^2$	ΔR^2	B	SE	β	t
Step 1	.303	.092	.016					
Tokens (I)					-0.001	0.002	-.052	-0.253
Total MRQ					0.339	0.221	.317	1.533
Step 2	.651	.423	.319	.332**				
Tokens (I)					0.000	0.002	-.021	-.112
Total MRQ					0.500	0.193	.467	2.589*
Verbal WM					-0.335	0.110	-.586	-3.032**
Visuospatial WM					-0.010	0.081	-.025	-0.123

*p < .05; **p < .01; ***p < .001

In the original analyses, spontaneous gesturers were compared to non-spontaneous gesturers in terms of their reading comprehension outcomes. In order to preserve the variability in gesture rate, this distinction was not used in the current analyses. In the current analyses, iconic gesture rate was used as a continuous variable. Additionally, hierarchical multiple regression was used rather than stepwise regression. Furthermore, intermittent summary and motivation were statistically controlled in step 1 of the analyses. This contrasts to the original analyses where these variables were entered after first assessing the influence of gesture production upon reading comprehension outcomes.

Hypothesis 6: *Individuals who do not gesture spontaneously will perform better on Low Inference assessments than individuals who gesture spontaneously*

In order to determine whether gesture production in the Control condition influenced Low and High inference measures of reading comprehension, hierarchical multiple regression analyses were conducted. Since intermittent summary length and motivation differed significantly across spontaneous gesturers and non-gesturers, they were statistically controlled in step 1 of the analyses. In step 2, iconic gesture rate was entered.

Low Inference

In the first step, intermittent summary length and motivation were entered into the regression. This model was statistically significant F (2, 24) = 5.562, $p < .05$ and explained 26.0% of the variance in Low Inference scores. In the second step, iconic gesture rate was entered into the multiple regression: This model was statistically significant F (3, 23) = 3.557, $p < 0.05$ and explained 22.8% of the variance in Low Inference scores (see Table 12).

Table 12

Regression model predicting Low Inference outcomes in the Control condition

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.563	.317	.260					
Tokens (I)					0.036	0.018	.349	1.951†
Total MRQ					3.266	1.723	.339	1.896†
Step 2	.563	.317	.228	.000				
Tokens (I)					0.036	0.019	.350	1.912†
Total MRQ					3.223	1.844	.335	1.748
Iconic rate					0.126	1.626	.014	0.078

†p<0.07 ; *p < .05; **p < .01; ***p < .001

Hypothesis 7: *Individuals who gesture spontaneously will perform better on High Inference assessments than individuals who do not gesture spontaneously*

In order to determine whether gesture production in the Control condition influenced High Inference measures of reading comprehension, hierarchical multiple regression analyses were conducted. Since intermittent summary length and motivation differed significantly across spontaneous gesturers and non-gesturers, they were statistically controlled in step 1 of the analyses. In step 2, iconic gesture rate was entered.

High Inference

In the first step, intermittent summary length and motivation were entered into the regression. This model was statistically significant $F\ (2, 24) = 30.998$, $p < .001$ and explained 69.8% of the variance in High Inference scores. In the second step, iconic gesture rate was entered into the multiple regression: This model was statistically significant $F\ (3, 23) = 20.635$, $p < 0.001$ and explained 69.4% of the variance in High Inference scores (see Table 13).

Table 13

Regression model predicting High Inference outcomes among the gesturers and non-gesturers

	R	R^2	*Adj.* R^2	ΔR^2	*B*	*SE*	β	*t*
Step 1	.849	.721	.698					
Tokens (I)					0.093	0.013	.851	7.434***
Total MRQ					-0.051	1.179	-.005	-0.043
Step 2	.854	.729	.694	.008				
Tokens (I)					0.093	0.013	.846	7.334***
Total MRQ					0.259	1.243	.025	0.208
Iconic rate					-0.914	1.096	-.095	-0.834

†p<0.07 ; **p* < .05; ***p* < .01; ****p* < .001

Summary of results from Part 3: The original analyses revealed that iconic gesture rate was significantly predicted by motivation and verbal working memory. This was replicated using an alternative manner of analysis. The original analyses demonstrated that participants who gestured tended to show advantages on both Low and High Inference measures of reading comprehension, but that these advantages were no longer significant when controlling for summary length. It is not surprising then that when initially controlling for summary length, gesture production is not a significant predictor of reading comprehension outcomes.

Implications

When comparing the results from the original analyses used with the results form the current analyses, it is evident that there are consistencies and differences. In terms of consistencies, both methods of analysis tend to highlight motivation as a critical predictor of reading comprehension among participants in the Control and Gesture conditions. Another consistency is that participants in the Written condition tend to excel on High Inference measures of reading comprehension. Yet another consistency is that iconic gesture rate is positively associated with motivation for reading, and negatively associated with verbal working memory. The most substantial difference that should be noted between the original and current analyses concerns the influence of spontaneous gesture production upon reading comprehension outcomes. This was likely influenced by the fact that iconic gesture production was assessed as a continuous variable in the current analyses, and by the fact that word tokens were statistically controlled before the influence of iconic gesture production was assessed. This is an important consideration since some researchers investigate iconic gesture production as a continuous variable (i.e. Kelly et al., 2011), while others analyze it as a dichotomous variable (i.e., Alibali et al., 2011). Additionally, narrative length is not always considered when assessing summary generation in relation to reading comprehension outcomes (i.e. Pecjak et al., 2011).

In conclusion, the information provided in this Appendix reveals that the specific manner in which the analyses are conducted can have an influential effect upon the evaluation of the usefulness of gesture production in the context of a reading comprehension task.