Working Memory and the Maintenance of Multiple Inferences

by

Allison K. Smith

Bachelors of Philosophy in Communication Science and Disorders, University of Pittsburgh

2016

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Bachelors of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

School of Health and Rehabilitation Sciences

This thesis was presented

by

Allison K. Smith

It was defended on

March, 30, 2016

and approved by

Erin Lundblom, PhD, Communication Science and Disorders, SHRS

Susan Shaiman, PhD, Communication Science and Disorders, SHRS

Heather Harris Wright, PhD, Communication Sciences and Disorders, East Carolina University

Thesis Director: Connie A. Tompkins, PhD, Department of Communication Science and Disorders, SHRS

Working Memory and the Maintenance of Multiple Inferences

Allison K. Smith, BPhil

University of Pittsburgh, 2016

The influence of working memory on facets of comprehension such as inferencing has been widely investigated. Working memory capacity for language (WMCL), a common measure of the central executive component of working memory, has been linked to inferencing and particularly to the maintenance or narrowing of inferences when multiple interpretations are generated. Conflicting conclusions are evident, however. Some work suggests that poor WMCL leads to overly narrowing inferences, attributing this performance to a lack of cognitive fuel to maintain simultaneous interpretations. Other work concludes that poor WMCL yields a difficulty in narrowing inferences to reach a single conclusion. And some data from the realm of syntax suggest that maintaining multiple options is a strength, available only to individuals with good WMCL. To further explore this issue, this experiment investigated inference narrowing in neurotypical adults with High vs. Low WMCL. In a thinking-outloud task, participants responded with whatever came to mind after hearing each sentence of 8-to-9 sentence narratives. Each response was coded as to whether it reflected a single inference, multiple inferences, or maintained inference.

This study also expanded the conceptual scope of prior investigations by assessing inference maintenance in relation to a newer component of the working memory system, the episodic buffer. The episodic buffer participates in combining and encoding the processes and outputs of other components of working memory and long-term memory. Its functioning was measured in terms of relative verbatim recall of a separate set of coherent versus scrambled stories.

One main result of this study was that the WMCL groups did not differ in the maintenance or narrowing of multiple inferences. Responses with multiple inferences declined for the Low WMCL group as the narratives progressed, but this result was not significant. Another major result was that the episodic buffer measure significantly predicted inference narrowing. Secondary analyses assessed potential differences between WMCL groups in inference revision and total inferences, neither of which were significant. Discussion centers around the limitations of relying solely on the central executive concept to predict inferencing outcomes, in light of its emphasis on the flexible allocation of cognitive resources, and suggests that a measure of episodic buffer functioning may be a better predictor, at least in some cases.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

I want to extend my sincerest gratitude and appreciation to those who have helped and supported me along this adventure. The completion of my thesis would not have been possible without the contributions from those listed below.

The University Honors College, for providing me with the opportunity to explore my idea in depth. It was through this project that I found my passion for research and communication disorders. The advising staff of the School of Health and Rehabilitation Sciences, Communication Science and Disorders Department, specifically Dr. Erin Lundblom. Thank you for introducing me to the BPhil, and encouraging me through every step of this process, I would not have gained the confidence and experiences in my undergraduate career without your guidance.

Fellow lab member, Tori Scharp, for answering my many questions and assisting me in the Tompkins Language Lab throughout the past two years. Also, I would like to extend my gratitude to the University of Pittsburgh's CTSI (Clinical and Translation Science Institute) registry [1] for including my study. Also, to my grandmother for spreading the word about my study in my hometown.

Dr. Margaret Blake, Dr. Jill Brady, and Dr. Connie Tompkins for providing me with materials from their studies. Dr. Erin Lundblom, Dr. Susan Shaiman, and Dr. Heather Harris Wright for taking time out of their very busy schedules to serve as my committee members, especially to Dr. Wright for traveling to the University of Pittsburgh for my defense.

The participants who volunteered for my study, I cannot thank enough. I am truly grateful that they took the time out of their schedules to assist me, and without their help and dedication this would not have been possible.

Dr. Connie Tompkins, for all of her support throughout the past two years. I cannot express how much I have learned and valued this experience with her, and I could not have asked for a better mentor. Thank you for the many meetings, draft revisions, and brainstorming sessions that we had. I hope to be able to use her constant guidance and dedication for the field of research and speech-language pathology as I purse my academic and profession career.

To my friends, thank you for listening to me and supporting me throughout these past two years. Especially, my roommates Taylor Hill and Haleigh Stapleton, for staying up many nights with me to practice my protocol and listen to my ideas. Also, my family for encouraging all of my academic dreams and never letting me give up.

[1] The project described was supported by the National Institutes of Health through Grant Number UL1TR000005.

1.0 INTRODUCTION

Each day individuals are faced with information that they must maintain and store for a short amount of time. Some examples include recalling a meeting time (Monday, at 10:30); remembering facts for an exam (water is made of Hydrogen and Oxygen); and considering multiple definitions of a word (row) in a given context (Henry, 2011). This critical information must be processed and stored for a sufficient amount of time for individuals to manipulate and mull over each aspect and determine what is relevant for the task at hand. The system that supports this temporary processing and storage capacity is known as working memory. Working memory is a system in which individuals temporarily store information as they perform thinking and reasoning tasks involved in language and cognitive processing (e.g., Baddeley, 2000; Just & Carpenter, 1992).

Working memory is a limited capacity cognitive system that underlies many domains of processing including reasoning, problem solving, and language comprehension (Just & Carpenter, 1992). Each individual has a different working memory capacity that he or she can allocate to simultaneously process and store information to succeed in these cognitive domains.

With regard to language comprehension, evidence indicates that working memory plays a critical role in initial processing, storage, and integration of spoken or written language in a conversation. Although this cognitive system is particularly important for language comprehension, there is some disagreement in the literature about how working memory may influence aspects of comprehension.

One example of this discrepancy comes from the domain of inferencing, which refers to the ability to make predictions or reach conclusions that go beyond what is directly stated. Specifically, Blake has proposed different thoughts about working memory and inference deficits. Blake and Lesniewicz (2005) examined older adult participants' generation of multiple inferences when presented with text scenarios that supported multiple possible outcomes. Later, the texts provided information that made many of those early inferences less likely, and the authors were interested in how well participants narrowed their possible inferences to reach the best conclusion. Based on a thinking outloud protocol, they found that participants with right hemisphere brain damage did not narrow their predictions. The authors attributed this performance to low overall working memory capacity for language (WMCL). From these results Blake and Lesniewicz concluded that narrowing requires efficient processes like integrating possible inferences with the rest of the context and selecting more appropriate inferences, processes that may require a large amount of WMCL. Therefore, they suggested, individuals with a low WMCL do not narrow but rather maintain numerous inferences. However, in another study Blake (2009) reported that adults with right hemisphere brain damage who had low WMCL overly narrowed their predictions, keeping only a few options open. She proposed that the participants had to purge some inferences from their working memory because the alternate inferences exceeded their capacity.

More generally, Just and Carpenter (1992) proposed that keeping more comprehension options available is not a consequence of low working memory. Rather, they provided evidence from the domain of syntax that individuals with high working memory can keep more options open. In work with college students, they demonstrated that the larger an individual's WMCL is, the more grammatical interpretations of a challenging sentence they maintain. The current study attempted to unravel the relationship between working memory and the maintenance or narrowing of multiple inferences.

1.1 THE BADDELEY MODEL OF WORKING MEMORY

One of the most influential models of working memory is that of Baddeley and colleagues, which lays out a framework of four components: a central executive, a phonological loop, a visuospatial sketchpad, and an episodic buffer (Baddeley, 2000). This model has undergone multiple revisions since its origination.

Much of the work on working memory and inferencing has equated working memory with the central executive in Baddeley's model (e.g., Blake, 2009; Blake & Lesniewicz, 2005; Just & Carpenter, 1992; Tompkins, Bloise, Timko, & Baumgaertner, 1994). Per the original model, the central executive was involved in allocating and directing working memory resources so that they are sufficient to accomplish a goal (Baddeley & Hitch, 1974). The central executive originally was attributed some storage capacity (Baddeley & Hitch, 1974) but was remodeled to be a processing-only component of working memory (Baddeley & Logie, 1999). Just and Carpenter (1992) described working memory as akin to the central executive, in enabling simultaneous processing and storage that trade-off in their demands on the amount of 'cognitive fuel' that each individual has available. The working memory task used by Blake (Blake &

Lesniewicz, 2005; Blake, 2009), Tompkins and colleagues' (1994) auditory working memory measure, was developed to capture that concept.

The visuospatial sketchpad in Baddeley's model is responsible for maintaining visual and spatial information. Because this study focused on auditory language processing, the visuospatial sketchpad will not be considered further. The phonological loop is responsible for storing and maintaining phonological information (Nobre, Rodrigues, Sbicigo, Piccolo, Zortea, Junior, & de Salles, 2013). The phonological loop interacts with the central executive through collaborations with the episodic buffer. The episodic buffer was added into Baddeley's model to act as a multidimensional storage system that integrates information from different sources of the central executive and long-term memory (Baddeley, 2002).

Although overlooked thus far in studies of working memory and inferencing, the episodic buffer may play an important role. The episodic buffer accounts for the temporary storage of information while also combining and encoding the processes and outputs of other components of the Baddeley model and long-term memory (Baddeley, 2012). The episodic buffer was added to the original model to allow the integration of the other three aspects of working memory to occur in a common storage area (Baddeley, 2012). Thus, through the episodic buffer, the phonological loop and visuospatial sketchpad are better coordinated with the central executive. Relevant to the current study, the episodic buffer is thought to be involved in processing sentences and prose (Christoffels, 2006).

It is unclear how to measure the workings of the episodic buffer (Baddeley, 2012), but a systematic review by Nobre et. al, (2013) suggested that one way to capture its contribution is to contrast recall of coherent vs. incoherent language. The idea is that the syntactic and semantic information in coherent language engages the episodic buffer in drawing on contributions from long-term memory to chunk that information. Chunking is defined as integrating several input items (e.g., words or numbers) into one representation (Nobre et al., 2013). Following from Baddeley's proposal (2002) that the episodic buffer is responsible for binding working memory processes and outputs with long-term memory, tasks that require verbal binding in the form of words or sentences provide a way to evaluate the functioning of the episodic buffer.

Jeffries, Lambon, and Baddeley (2004) developed such a task, comparing the recall of identical words that were integrated into related and unrelated sentences. In the related sentences, the stimulus words were arranged to form a coherent unit whereas in the unrelated sentences, the words when combined did not form a coherent whole. Results indicated that the stimulus words were easier to recall in related sentences than in unrelated sentences. The authors concluded that the contribution from long-term memory, mediated by the episodic buffer, aided in the recall of the related condition.

Christoffels (2006) tested the recall of incoherent and coherent stories, and found similarly to Jeffries et al. (2004) that her adult participants did better when sentences in a story had a meaningful relation to one another, forming a coherent structure, rather than when sentence arrangement was incoherent. She described this result as a coherent advantage effect, defined as better recall with a coherent story structure than for the same sentences presented in a different, incoherent, order. In a similar manner Kapikian and Briscoe (2012) used one set of sentences arranged into either coherent or incoherent stories. Results again indicated that the coherent stimuli were more easily recalled. The authors concluded, as Jeffries et al. (2004) and Christoffels did, that coherent stories relied on the episodic buffer to integrate long-term memory information.

1.2 THE ENGLE AND KANE MODEL OF WORKING MEMORY

Kane and Engle (2003) put forward a working memory capacity model that overlaps in a number of ways with the concepts explained above. The researchers also propose two mechanisms that maintain and occupy cognitive fuel: goal maintenance and response competition. These mechanisms are incorporated into their Two Factor Theory of Control (Engle and Kane, 2004). Goal maintenance can be defined as sustaining a goal in order to accurately complete a task at hand (Kane and Engle, 2003). Response competition occurs when a task prompts or promotes interfering response options, especially when one option is strongly predominant. Kane and Engle (2003) deduced that these two factors of control contribute to the nature of response and maintenance of information in interfering tasks. When interference is absent, active goal maintenance can be sustained, promoting accurate performance. However, when interference generates response competition and cognitive fuel is devoted to quickly performing the task rather than to selecting the most relevant response, more errors and slower response times are likely.

These concepts can possibly be connected to the issue of maintaining or narrowing inferences. In a think-out-loud protocol such as that used by Blake and Lesniewicz (2005), where the participants are instructed that the goal is to let the mind connect to anything in the scenarios, generating and maintaining multiple inferences could be evidence of good goal maintenance. As the scenarios begin to point toward one inference over another, narrowing the multiple inferences may bear some similarity to the process of resolving response competition.

1.3 STUDY AIMS AND OBJECTIVES

This study returned to the contrasting conclusions of Blake and Lesniewicz (2005), that groups of participants with high and low WMCL were equally likely to keep multiple inference options open, and Blake (2009), who found that individuals with high WMCL kept more options than those with low WMCL. Just and Carpenter (1992) also aligned high working memory (in the central executive sense) with a particular prowess in maintaining multiple options. Based on these contradictions, one primary aim of the current study was to assess whether individuals who differ in estimated WMCL (high vs. low) also differ in the maintenance of multiple inferences from text. One possible outcome, consistent with Blake and Lesniewicz (2005), is that adults with low WMCL are more likely than those with high WMCL to maintain multiple inferences in contexts that call for narrowing them. On the other hand, extending the evidence from Just and Carpenter (1992), it is possible that individuals with high WMCL are no less likely to maintain multiple inferences than those with low WMCL, and in fact may be better at doing so.

A second aim was to expand the scope of previous studies of the relationship between working memory and inferencing by taking into consideration the interaction of the episodic buffer with the central executive (Baddeley, 2000). Specifically, this study also assessed whether a measure of episodic buffer functioning would predict inferencing outcomes.

A secondary data analysis examined whether individuals with high WMCL reinterpret their initial inferences more often than those with low WMCL, given work suggesting that adults with better WMCL may be better at revising inferences than those with lower WMCL (e.g., Tompkins et al., 1994). Another secondary data analysis assessed whether individuals with high WMCL generate more inferences in general than those with low WMCL. Overall this study was designed to help to clarify the nature of the relationship between working memory and the maintenance of multiple inferences, when a text cues some of those inferences as being less likely.

2.0 METHODS

2.1 SUBJECTS

Twenty neurologically healthy adults completed the experiment, 10 with high WMCL and 10 with low WMCL. The WMCL groups were constituted with reference to the number of word recall errors they made on the Auditory Working Memory measure (Tompkins et al., 1994), described in Section 2.2.1 below. As evident in Table 1, participants in the High WMCL group made 4 or fewer errors on this task, whereas those in the Low WMCL group made 8 or more errors.

All participants provided voluntary consent prior to testing. All were between the ages of 49-84, learned only English when developing language as a child, and self-reported no history of or current abuse of substances or alcohol. The subjects' other characteristics are listed in Table 1. The two subject groups differed in the measure of central executive functioning, which is the number of word recall errors on the Auditory Working Memory task (Tompkins et al., 1994) (t (11.9)[b] = -8.99; $p < .01$). There were no group differences in the other variables in the table (all t (18) < /1.04/; all $p > 0.31$).

---

[b] Degrees of freedom adjusted for unequal variances

Table 1. Characteristics of two subject groups

Characteristics	High WMCL	Low WMCL
Auditory Working Memory[a]		
Word Recall Errors		
Mean (Std. Dev.)	2.9 (0.1)	10.3 (2.4)
Range	1-4	8-15
True/ False Errors		
Mean (Std. Dev.)	0.1 (0.3)	0.1 (0.3)
Range	0-1	0-1
Sex	6 Females, 4 Males	8 Females, 2 Males
Age		
Mean (Std. Dev.)	62.0 (9.4)	66.6 (10.4)
Range	51-81	49-84
Education (in years)		
Mean (Std. Dev.)	16.0 (3.3)	15.8 (2.4)
Range	12-22	12-20
Peabody Picture Vocabulary Test-Revised[b]		
Mean (Std. Dev.)	166.8 (4.4)	165.0 (4.2)
Range	158-173	158-171

Note. WMCL = Working Memory Capacity for Language; Std. Dev. = Standard Deviation.

[a] Tompkins, C.A., Bloise, C. G. R., Timko, M. L., & Baumgaertner, A. (1994). Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech and Hearing Research*, 37, 896-912.

[b] Dunn, L. M., & Dunn, L.M. (2000). Peabody Picture Vocabulary Test III. Circle Pines, MN: American Guidance Service.

Subjects were recruited through three avenues. The first was the Tompkins Language Laboratory Research Registry, comprising individuals who previously provided voluntary consent to be a part of a registry for participation in studies by the Tompkins Language Lab. All of these participants were screened for subjective reports of neurological status and self-reported alcohol and drug abuse. The second method was to recruit through the University of Pittsburgh's Clinical Translational Scientific Institute (CTSI). Individuals recruited from CTSI had provided voluntary consent to be a part of the registry for participation in studies that qualify with their ICD-9 codes. They were referred to the primary investigator (AKS) when they met several basic criteria concerning age, neurological status, and self-reported alcohol and drug abuse. The third method used for recruitment was word-of-mouth, by posting or distributing a flyer in the community. Interested individuals contacted the primary investigator who then screened them and obtained voluntary consent.

During an initial telephone screening the investigator explained the procedure to all potential participants, and then asked eligibility questions to determine whether they could participate in the experiment. The questions asked about neurological, sensory, and substance abuse status located in Appendix A. In order for the subject to be eligible, he or she had to answer "No" to the questions about neurological status and substance abuse. If the subject remained eligible after this initial telephone screening, informed consent was obtained for the current study.

At the beginning of the experimental testing session, hearing and dementia screenings were conducted. There were two ways to pass the hearing screening. First, subjects passed if they achieved a pure-tone average of less than 35 dB HL at 500 Hz, 1000 Hz, and 2000 Hz. If the pure-tone average of 35 dB HL was not achieved, a behavioral speech-recognition and repetition hearing screening was administered. For this screening, the participants were asked to repeat 10 one or two syllable words, loaded with fricative consonants, read by the primary investigator. The investigator held a clipboard in front of her mouth to prevent subjects from having a lip-reading advantage. To pass, 11/12 or 91.7% accuracy needed to be achieved. This procedure was used with three participants. Dementia was ruled out using the Mini Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975). Eligible participants had to achieve a minimum score of 27.

2.2 STUDY OVERVIEW AND DESIGN

Participants in this study completed three primary tasks: (1) an auditory working memory task to estimate WMCL and assign WMCL group, (2) a thinking outloud inferencing task to quantify the generation, maintenance, and narrowing of multiple inferences, and (3) a coherent and scrambled story task to evaluate the functioning of the episodic buffer. Each of these tasks was presented auditorily. The study used a descriptive, two factor mixed design with participant group (Low WMCL, High WMCL) as the between-subject factor and task condition (as elaborated below) as a within-subject factor.

2.2.1 Auditory Working Memory task

WMCL was measured using the Auditory Working Memory task of Tompkins et al (1994), designed to evaluate a listener's capacity for simultaneous processing and storage of language input. In this task, participants listen to 12 sets of short simple active declarative sentences that represent common knowledge (e.g., You sit on a chair) or counterfactuals (e.g., Trains can fly). There are 42 sentences total, arranged in sets that increase in size from two sentences per set to five sentences per set. Immediately after hearing each sentence, participants indicate whether it is true or false while remembering the last word of the sentence for spoken recall at the end of each sentence set. The stimuli aim to stimulate both concurrent processing and information storage, and were designed so that the True/False judgment could not be made until the final word (Tompkins et. al, 1994). A sample of the stimuli is located in Appendix B.

As previously noted, participants were divided into High vs. Low WMCL groups based on the primary outcome measure for this task: number of word recall errors. The number of True/False errors is a secondary outcome of this task. Because the sentence stimuli tap highly common knowledge, these errors appear to reflect momentary distractions and are always extremely infrequent, even for brain-damaged individuals (e.g., Tompkins et al., 1994).

2.2.2 Inferencing task

The inferencing task for this study was the thinking outloud task used by Blake and Lesniewicz (2005), who graciously provided their stimuli. The eight stimulus stories in this task describe one or two main characters engaging in everyday scenarios (e.g., doing errands, working as a waitress). There are 3 versions of each story. Two versions, the experimental stories, were constructed to suggest a negative outcome with either a high or low degree of predictability. The third version is a control stimulus, which is either entirely positive or contains only a minor irritant.

To keep the length of the task manageable, this study used only the low predictability stories as experimental inference-eliciting stimuli. The low predictability stimulus stories describe scenarios (e.g. a waitress dealing with an unhappy customer) that skew toward predicting a negative outcome (e.g., she will dump the soup on his head) and then suggest an inference revision when that negative outcome does not occur (e.g. she takes the soup back to the kitchen). The control stories were administered as well, to disguise the nature of the experimental stimuli. One of the original stories was excluded because of discomfort about presenting the content to adult participants who were significantly older than the PI. Therefore the participants in this study listened to 14 stories, seven experimental and seven control. Each was eight or nine sentences in length. One other minor modification was made to one stimulus story. Specifically, the phrase “first encounter” was changed to “friendship” in order to provide better clarity. Appendix C provides several complete stimulus examples.

2.2.3 Episodic buffer task

The task to assess the contributions of the episodic buffer was the coherent and incoherent (scrambled) story task from Brady (2009). Participants in this study listened to seven coherent and seven scrambled trials. In the Coherent condition the sentences were arranged into a clear and comprehensible story. That story introduced a character who was presented with a statement or situation and expected to react with an opinion or with a statement of belief or knowledge[c]. The Scrambled condition stimuli began and ended with the first and last sentences from the Coherent condition stimuli. However, the intervening six sentences were quasi-randomly selected from different coherent texts, with no more than one sentence from the same text. This was done to reduce the likelihood that the comprehender would generate a global representation of the stimulus. For Brady’s purposes, the verbs in each sentence were chosen to have no more than one predominant meaning, to be one-to-three syllables in length, and to subcategorize for both noun phrases and sentential complements. A sample stimulus from each condition is located in Appendix D.

[c] The Coherent stimuli were from Brady (2009) 2S condition, i.e., the condition in which two of the eight verbs resolved to sentential complements.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Task Construction

Auditory working memory task. This study used the original stimulus recording from the Tompkins et al. (1994) study, and it was converted to a .wav file for digital presentation to participants. In the original recording, a female produced the stimuli at a slow normal rate, approximately 180 words per minute (Tompkins et al, 1994). No particular emphasis was used on any word. The beginning of each set of sentences was signaled with the word “ready” and a one second pause. After each sentence there was a three second pause, and between each set there was a five second pause. The task was presented in a single block.

Inferencing task. Stimuli for this task were recorded by the PI (AKS) in a quiet environment using GarageBand for Apple MacBook Pro 2011. The stimuli were produced at a slow normal rate of approximately 120 words per minute with a two second pause after each sentence indicating that the sentence was complete. The mean length of the stories was 58 seconds (standard deviation = 5.5 seconds). The stimuli for this task were split into two blocks. Each block contained 3 or 4 experimental or control story stimuli, pseudo-randomly arranged so that no more than 2 of each stimulus type occurred in sequence.

Episodic buffer task. Stimulus recording was done as it was for the Inferencing task. The average length of the stories was 57 seconds (standard deviation = 3.6 seconds). These stimuli were also divided into two blocks following the same procedures as those for the Inferencing task.

2.3.2 Experimental Apparatus and Procedures

Stimuli were played on the PI's portable laptop computer, MacBook Pro late 2011 version, and presented through Sony Over Ear headphones, model, MDRZX110/BLK at a comfortable loudness level as selected by the participant. Each participant took part in one session of approximately one hour and thirty minutes. The sessions took place at the subject's home, a public library (one participant), or the University of Pittsburgh Adult Cognitive Language Lab.

The session began with the screening measures described above, followed by the WMCL task and a playlist for the inferencing and episodic buffer tasks. The order of inferencing and episodic buffer tasks was counterbalanced across participants. Within each task, the order of presentation of Blocks 1 and 2 was also counterbalanced. The two blocks for each task were separated by 10 items from the PPVT-R (Dunn & Dunn, 2000), and the transition from one task to the other was also separated by 10 items from the vocabulary test. During testing, the examiner sat back from the subject and did not interact until an experimental block was completed. Then only general encouragement was given if needed.

2.3.3 Procedures for each task

Auditory working memory task. Participants were informed that they would hear several sets of sentences and would be asked to do two things. First, they would indicate whether each sentence was True or False, and second, they would remember the final word of each sentence for recall when requested by the examiner. Two response cards were created, one labeled *True* and the other *False*. These response labels were typed in the center of plain white index cards, in all capital letters using a bolded, 70-point Times New Roman font. During the testing session, these response cards were aligned horizontally on the testing surface at the participant's midline. Participants indicated their True/False response after each sentence by pointing to the correct index card. At the end of each set of sentences, the investigator paused and pointed to the participant to cue recall of the last word of each sentence in the set. The order of recall for sentence-final words was not considered in scoring. The PI introduced the task with two spoken examples. Scoring took place during the session.

Inferencing task. Participants were instructed that they would hear a set of stories, and asked to respond after each sentence of each story with their predictions of what would happen in the story, any other ideas that come to their mind as they listened, or if anything seemed unusual. The exact instructions are included in Appendix E. Participants responded after the PI paused the stimulus recording and pointed at the participant. The responses to the task were audiorecorded on an iPad 2 using the Recorder application for later orthographic transcription and coding.

Episodic buffer task. Again, participants were told that they would hear a set of stories. This time they were instructed to recall each stimulus outloud as precisely as they could. Specific instructions are contained in Appendix E. After each item, the PI paused the playlist and pointed to the participant. Participants’ responses were recorded in the same manner as those for the inferencing task, for later transcription and scoring.

2.4 DATA PREPARATION

2.4.1 Transcription

Participants’ spoken responses were orthographically transcribed by the PI, who subsequently cross-checked each transcription for accuracy. For the inferencing task, inter-rater reliability assessment (AKS versus CT) and intra-rater reliability assessment (AKS time 1 versus time 2) was performed on 20% of the experimental stimuli, which were selected to demonstrate varied lengths and complexities of responses across participants. When calculating transcription agreements and disagreements, differences in the following were excluded from consideration: fillers (e.g. um), comments about the task (e.g. that was confusing), contracted or shortened forms versus noncontracted forms (e.g. wanna versus want to), and false starts (e.g. The, the man). Total inter-rater reliability averaged 98% (97-99). Total intra-rater reliability averaged 97% (95-99). In light of these strong results, transcription reliability was not assessed separately for the responses to the episodic buffer task.

2.4.2 Scoring

Inferencing Task. Following transcription of responses to the experimental Inferencing task stimuli, the response(s) to each sentence were coded using four main categories: Inference, Multiple Inference, Maintained Local Inference, and Maintained Distant Inference. An inference (I) was coded when the participant generated an original interpretation or prediction that had not been explicitly stated previously in the stimulus. Multiple Inference (MI) was coded when the participant made two or more inferences in response to one sentence (e.g., He is going to go to the counter and pay for the ring). Maintained Local inference (ML) was coded for a response that repeated a previous inference, generated less than three sentences prior to the current response. Maintained Distant Inference (MD) was coded when a prediction or theme was reinstated from three or more sentences prior. There were multiple possible codes for each response, because the response to a single stimulus sentence could range from a single short phrase or sentence to multiple elaborated sentences. To equate response opportunities across stories that differed in length, we did not code the response to the last sentence for the nine-sentence stories (N=6). Other types of responses (e.g., comment on the task, cliché, no response) were noted but excluded from data analysis. Some additional rules for coding the main categories are contained in Table 2.

Table 2. Inference task: Coding rules

Inference	This code applies when there is one original inference – an interpretation that was not explicitly stated in the story - made by the participant in his or her response to a stimulus sentence. Synonymous interpretations are counted as a single inference (i.e. peaceful and quiet)
Multiple Inference	Multiple Inferences are coded within a single response when there are two or more inferences within that response. The inferences must be original/new, and not occurring in previous responses.
Maintained Local Inference	Coded when the participant responds with a recurring theme or prediction they originally mentioned no more than two stimulus sentences previously. Maintained Local can co-occur with Multiple Inference if one inference is novel and the other is maintained from a previous response according to these criteria.
Maintained Distant Inference	Coded when a response returns to an inference theme that was originally made in response to a stimulus sentence at least three sentences prior. Maintained Distance can co-occur with Multiple Inference if one inference is novel and the other is maintained from a previous response according to these criteria.
No Code	If the response is a cliché (e.g., happily ever after), comment on the task (e.g., that made no sense), or no response (e.g., no comment, next one), none of the major codes are applied.

The same 20% of experimental Inferencing task transcripts that were used to assess transcription reliability were also assessed for inter- and intra-rater reliability of the inference codes. The results are provided in Table 3.

Table 3. Inter-rater and intra-rater reliability: Percent agreement for inferencing task codes

	Inter-rater	Intra-rater
Inference		
Mean	95.3%	94.1%
Range	91% - 97%	93% - 96%
Multiple Inference		
Mean	92.8%	90.8%
Range	83% - 100%	85% - 97%
Maintained Local Inference		
Mean	74.8%	100 %
Range	33%[a] -100%	100 % - 100%
Maintained Distant Inference		
Mean	91.3%	92.8 %
Range	80% - 100%	80% - 100%

[a] In the subset of transcripts assessed for reliability, the PI coded 1 instance of Maintained Local Inference and the second rater coded 3, thus the low agreement here.

Episodic Buffer Task. For the episodic buffer task, the mean length of the coherent story retells was 57 words (S.D. = 29 words), and the mean length of the scrambled story retells was 32 words (S.D. = 21 words). Transcripts were scored for verbatim recall accuracy. This was operationalized as the percentage of words recalled in any order (after Jeffries et al., 2004).

2.5 OUTCOME MEASURES

Two primary outcomes were chosen to capture potential WMCL differences in maintaining multiple inferences versus narrowing inferences. The first was based on changes in the number of MI responses as the stimulus texts progressed. We calculated the number of responses receiving this code for the first and second half of each stimulus (responses to sentences 1-4 vs. sentences 5-8). If MI codes decreased from the first to the second half of the stimuli, this would suggest that participants were narrowing their options. The second dependent measure was the number of responses coded as MD in the second half of each stimulus, to index, as another sign of narrowing, the degree to which participants were returning to prior inferences.

Inference revisions were also coded as a potentially interesting secondary measure, given work suggesting that adults with better WMCL may be better at revising inferences in some conditions than are those with lower WMCL (e.g., Tompkins et al., 1994). Inference revisions were defined when a subsequent response explicitly rejected or reversed a prior I or MI response (e.g. “she was worried” changed to “she felt good”). The total number of inferences (I + MI + ML + MD) was also calculated.

The outcome measure for the episodic buffer task was relative verbatim recall accuracy for Coherent vs. Scrambled stimuli, expressed as a ratio (% recall Scrambled / % recall Coherent) and averaged across the 7 pairs of stimuli. A smaller value indicates a greater reliance on or proficiency of the episodic buffer.

3.0 RESULTS

3.1 PRELIMINARY ANALYSES

The PI's data were used for all analyses. One set of preliminary analyses assessed data distributions. First, all distributions were analyzed for outliers (defined as values +/- 3 standard deviations from the group mean). There were no outliers in the primary dependent measures of inferencing or in the episodic buffer task. Second, ratios of skewness and kurtosis to their respective standard deviations were examined to check assumptions of normality. The assumption was violated for one measure of MI (that for sentences 1-4); thus, nonparametric statistics were used for major analyses involving this variable. Third, the probability plots of the distributions of the measures to be correlated, e.g., the episodic buffer ratio, were inspected and all were found to represent continuous linear distributions that were evenly spaced.

A second set of preliminary analyses was conducted to examine the influence of sex differences on performance in the Inferencing and episodic buffer measures, with none found to be significant (all t (18) < 1.269; p > .11). Finally, a correlational analysis was performed to assess the relationship between vocabulary knowledge and the Inferencing and episodic buffer measures. Vocabulary knowledge, assessed by the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 2000), did not correlate significantly with any of the outcome measures (all r (18) < .384; p > .10).

3.2 PRIMARY ANALYSES

Table 4 presents the descriptive data for the Inferencing and Episodic Buffer variables, separately for the High and Low WMCL groups.

Table 4. Descriptive data for the experimental measures for two WMCL groups

Variables	High WMCL	Low WMCL
Multiple Inferences 1 (Responses to sentences 1-4)		
Mean (Std. Dev.)	10. 8 (12.7)	15.0 (10.4)
Range	0 - 45	0 – 36
Multiple Inferences 2 (Responses to sentences 5-8)		
Mean (Std. Dev.)	8.2 (7.3)	9.0 (9.2)
Range	0 - 25	0 – 30
Maintained Distant Inferences (Responses to sentences 5-8)		
Mean (Std. Dev.)	8.2 (4.2)	10.7 (5.4)
Range	3 - 16	1 – 22
Total Inference Revisions		
Mean (Std. Dev.)	4.9 (3.6)	4.0 (1.4)
Range	0 - 10	1 – 6
Total Inferences		
Mean (Std. Dev.)	63.3 (20.2)	68.6 (20.0)

Range	29 - 109	37 – 96
Coherent Stimuli Recall		
Mean (Std. Dev.)	23% (7%)	24% (12%)
Range	10 % – 30 %	10% – 50 %
Scrambled Stimuli Recall		
Mean (Std. Dev.)	14% (3%)	14% (8%)
Range	10% – 20%	4% – 40 %
Episodic Buffer Ratio		
Mean (Std. Dev.)	0.7 (0.9)	0.6 (0.9)
Range	0.5 – 0.9	0.4 – 0.8

Note. WMCL = Working Memory Capacity for Language; Std. Dev. = Standard Deviation.

One primary aim of the current study was to assess whether individuals who differ in estimated WMCL (High vs. Low) also differ in the maintenance of multiple inferences from text. The first analysis compared WMCL groups on the number of responses in the first and second half of each stimulus that were coded as MI (responses to sentences 1-4 vs. sentences 5-8). One assessment of the narrowing of inferences (Multiple Inference 1 vs. 2) was analyzed within group, separately for High and Low WMCL, with a nonparametric approach. There was no significant difference in MI 1 versus MI 2 for either the High or Low WMCL group, per Wilcoxon related-samples signed rank test (p = .175 and p = .074, respectively). For the between group analyses, the differences between WMCL groups for MI 1 and 2 were analyzed separately via Independent Samples Mann-Whitney U test. Neither result was significant (p = .143 and .971, respectively). Finally, an independent t-test was used to assess WMCL group differences in

MD inferences (sentences 5-8), which indicated narrowing of options by returning to prior inferences. Again, there was no significant difference (t (18) = 11.15; p =. 264).

A second aim was to expand the scope of previous studies of the relationship between working memory and inferencing by taking into consideration the interaction of the episodic buffer with the central executive (Baddeley, 2000). Nonparametric Spearman rank order correlations were run to assess the relationship between the ratio measure of episodic buffer functioning and the MI 1 and 2 variables. The correlation with MI 1 (responses to sentences 1-4) was non-significant (*rho* (18) = .25; p = .288), but the episodic buffer ratio did predict MI 2 (responses to sentence 5-8) (*rho* (18) = .473; p = .05). Smaller ratios were associated with fewer MI responses to the last half of the experimental inferencing stimuli. An assessment of potential WMCL group differences in episodic buffer functioning was nonsignificant (t (18) = .427, p = .95). Further, a Pearson correlation between episodic buffer ratio and MD inferences was nonsignificant (r (18) = .021, p = .93). The coherent advantage effect reported in other literature (e.g. Christoffels, 2006, Kapikian & Briscoe, 2012) was assessed by paired t-test for all participants combined, and found to be significant (t (19) = 8.85; p = .000).

3.3 SECONDARY ANALYSES

One secondary data analysis examined whether individuals with High WMCL reinterpret their initial inferences more often than those with Low WMCL, given work suggesting that adults with better WMCL may be better at revising inferences than those with lower WMCL (e.g., Tompkins et al., 1994). WMCL group differences in total number of Inference Revisions were assessed by independent *t*-test, and found to be nonsignificant (*t* (11.71)[d] = .735; *p* = .477).

Another secondary data analysis assessed whether individuals with better WMCL generate more inferences in general than those with poorer WMCL. Each WMCL group's total inferences were summed (Inferences + Multiple Inferences + Maintained Local Inferences + Maintained Distant Inferences). Although the Low WMCL group had a greater mean of total inferences than the High WMCL group, this difference was not statistically significant (*t* (18) = -0.59; *p* =. 562.

---

[d] Degrees of freedom adjusted for unequal variances

4.0 DISCUSSION

This study aimed to help to clarify the nature of the relationship between working memory and the maintenance of multiple inferences, as later parts of a stimulus text cued an earlier inference as less likely. The investigation was motivated in part by the contrasting conclusions of Blake and Lesniewicz (2005) and Blake (2009), and the Just and Carpenter (1992) evidence that high working memory (in the sense akin to the central executive) allows an individual to keep more options open. This study also expanded prior investigations of inferencing and working memory, which have mainly focused on the central executive component, by assessing the contribution of the episodic buffer (Baddeley, 2000). The results met with expectations in some regards, and not in others.

4.1 WMCL GROUP AND MULTIPLE INFERENCES

One primary aim of the current study was to assess whether individuals who differ in estimated WMCL (High vs. Low) also differ in the maintenance of multiple inferences from text. The first outcome measure for this aim (MI 1 versus 2) was chosen to assess whether there was a reduction in MI responses as the stimulus texts progressed. The differences between MI 1 and MI 2 were found to be nonsignificant. It is evident, though, that there was a big numerical drop from MI 1 to MI 2 for the Low WMCL group. The fact that this difference was not significant could be due in part to the large within-group variability and small sample size, which may have swamped a real difference. The lack of differences between the two WMCL groups also may have reflected this large within-group variability. The large overlap in frequency of MI responses between groups most likely reduced the power to detect a difference. Some participants in each group appeared to narrow their inferences while others did not. Another possible reason for the nonsignificant difference in MI 1 vs. MI 2, particularly for the Low WMCL group, is that power was further reduced by the application of more conservative nonparametric analyses. The nonparametric approach had to be used because the data distributions did not meet assumptions for parametric analysis.

Overall, the results for the Low WMCL group may seem compatible with the Blake and Lesniewicz (2005) observation that individuals with poor WMCL do not narrow inferences, and contrary to Blake's (2009) contention that people with poor WMCL overly narrow inferences. However, as just noted, the Low WMCL group in this study did evidence a numerical decline in the number of MI responses they made as the stimuli approached an end. In addition, it is not clear that the absence of narrowing for the Low WMCL group reflects a dearth of cognitive fuel, as Blake and Lesniewicz hypothesized, given the lack of WMCL group differences in inference narrowing in the current study.

The second primary dependent measure was MD 2, which signaled inference narrowing by a return to prior inferences. There were no significant WMCL group differences in this measure either, even though more liberal parametric statistics were used for the analysis.

Complementing the findings of lack of narrowing for the Low WMCL group, the High WMCL group did not show evidence of a superior ability to keep options open, per Just and

Carpenter (1992), despite their presumed abundance of cognitive fuel. The main conclusion from the aggregate of these results seems to be that the WMCL measure of central executive functioning may not be a pertinent measure for identifying inferencing outcomes, and particularly inference narrowing.

Inherently, the concept of flexible allocation of resources that characterizes the Central Executive (e.g., Just and Carpenter, 1992) makes it difficult to predict how higher or lower WMCL will affect resource-demanding performance. Different participants will have different priorities. Some individuals with limited WMCL may choose to allocate their cognitive fuel to generating and/or maintaining multiple inferences (e.g. she learned to live, and she adapted to life without her husband), while others many choose to allocate their resources to narrowing quickly and deciding on one outcome (e.g. he will buy the ring). For different studies of the relationship between WMCL and multiple inferencing, the differences in participant samples may produce average group results that lean one way or the other, generating conflicting conclusions. In future work, it may be possible to manipulate task instructions to control more precisely how participants choose to allocate their cognitive resources.

The concept of the Central Executive is also difficult to measure because of other confounding factors. An example can be seen in the responses of the four widows who participated. These women resonated particularly with the stories involving a widow, as evident in their off-task comments comparing their own situations to those in the stories. Implicit or explicit biases and priming of this sort are likely to affect participants' priorities and consequent allocation of cognitive fuel. In future work, larger samples would help to minimize the influence of individual biases like these. Despite this resonance with individual themes in the current study, the widows' priorities likely had no systematic effect on the results. Potential problems in this regard were minimized because there were two widows in each WMCL group. In addition, these four participants received different counterbalanced orders of stimulus presentation, so that hearing one version (experimental or control) first would not have systematically affected the response to the experimental stimuli.

4.2 EPISODIC BUFFER AND MULTIPLE INFERENCES

A second aim of this investigation was to expand the scope of previous studies of the relationship between working memory and inferencing by taking into consideration the interaction of the episodic buffer with the central executive (Baddeley, 2000). As in previous literature (e.g. Christoffels, 2006, Jeffries et. al, 2004, Kapikian & Briscoe, 2012, Nobre et. al, 2013), the coherent stimuli were recalled more accurately than the scrambled stimuli. Thus the current study further replicated the findings of a coherent advantage effect (e.g. Christoffels, 2006, Jeffries et. al, 2004, Kapikian & Briscoe, 2012): better recall of information within a coherent story structure than for the same information presented in a different, incoherent order.

These findings are consistent with the idea that chunking, defined as integration of several outputs into one representation, is occurring during the episodic buffer's interaction with the central executive. This concept can be linked to Baddeley's (2002) proposal that the episodic buffer integrates working memory processes with long-term memory. Further, the significant results from this study suggest that the contribution from long-term memory, aided by the episodic buffer, resulted in greater recall of the coherent structured stories.

The results of the current study found the measure of the episodic buffer, relative verbatim recall of scrambled versus coherent stories, to predict the extent to which inference narrowing occurred as the texts progressed. Better episodic buffer functioning was associated with a greater reduction in multiple inferences as individuals processed the information that was presented to them. This suggests that an episodic buffer measure could be a better predictor than WMCL for assessing certain inferencing outcomes. As a temporary storage and integrator with other working memory processes and long-term memory, the episodic buffer provides a common storage area for new representations (Baddeley, 2012).

4.3 ENGLE AND KANE MODEL

In relation to the Kane and Engle (2003) working memory capacity model, the results in this study can be considered in another light. First, the results suggest the WMCL groups were similar in goal maintenance (Kane & Engle, 2003). The express goal of the task, cued by the task instructions, was to produce multiple inferences and the groups did not differ on this factor. Also per the results, the subjects did not differ in narrowing as the scenarios began to point to one inference over another. This may suggest that the two WMCL groups have some similarities in the process of resolving response competition (Kane & Engle, 2003). However, this comparison may not be apt, as the Engle and Kane model was based on the Stroop task (Stroop, 1935), in which response competition takes the form of a strong habitual (prepotent) response that interferes with the task goal. The competition between inferences in the texts in this study was not of that same form. With this caveat in mind, individuals with better episodic functioning may be better at resolving some aspects of response competition in a narrative context.

4.4 SECONDARY AIMS

Further analysis was done to assess whether the frequency of inference revisions differed between the High and Low WMCL groups. Once again there were no significant group differences. To try to explain this result we return first to some of Just and Carpenter's (1992) theorizing. Consistent with their view, Tompkins and colleagues (1994) provide evidence that WMCL influences performance only when the task or process in question taxes an individual's capacity. Thus, in this study, it is possible that the inference revision process in which participants engaged was not challenging enough to meet or exceed the Low WMCL group's capabilities. However, many of the inference revisions were such that the initial inference was generated more than once before it was revised, similar to one of the conditions in Tompkins et al. (1994) in which inference revision was in fact linked to WMCL. The range of WMCL in this study, and mean across individuals in the two WMCL groups, was also similar to that for Tompkins et al. (1994) control group. Thus the difference in results between the two studies must lie in other factors. One other difference between the two studies is that in Tompkins et al., the inference to be revised was a mood or emotion that was strongly indicated, prompted, or cued. Thus, the challenges of inference revision could have been higher in Tompkins et al. due to the emotional content and/or the strength of the initial inference. The differences between the two studies may also reflect the heterogeneity of the Low WMCL group, whose WMCL performance ranged from 8-15 errors.

Finally, the total number of inferences generated throughout the task was assessed to see if there was a significant WMCL group difference. Although the Low WMCL group generated numerically more inferences on average, there was no significance difference between the two WMCL groups. This variable was investigated primarily to help shed light on any group differences that might emerge for particular aspects of inferencing in this study. Because no such differences occurred, total number of inferences will not be considered further.

4.5 LIMITATIONS

The primary focus of the current study was to clarify the nature of the relationship between working memory (WMCL and episodic buffer components) and the maintenance or narrowing of multiple inferences. As in all investigations, there are some limitations to consider.

In regards to subject participation, the sample size was small (N = 20, 10 per WMCL group). With a larger sample, inference narrowing could have been examined within WMCL group, to get a handle on the nature and consequences of the large within-group variability. This may be particularly important within the Low WMCL group, which could actually be more than one group (relatively more or less poor WMCL) because of the large range of possible errors that can occur (8-42 hypothetically; 8-15 in this study). In a preliminary assessment of the possible influence of within-WMCL-group variability, a measure of inference narrowing (calculated as

MI 1 – MI 2) was correlated with WMCL within the Low WMCL group. The result was nonsignificant (*rho* (9) = .569, p > .53), although this is not particularly surprising given the small sample and low-power, nonparametric test. In addition, however, heterogeneity on the WMCL measure is not the only factor responsible for lack of narrowing, because the High WMCL group is much more homogenous (with results ranging from 0-4 errors). A larger participant sample would also help to reduce concerns relating to individual biases or priorities, such as those that may have affected the responses of the widows as mentioned in a previous section. A final issue related to the participant sample is that the WMCL groups may not have been distinct enough to find any differences in inferencing. It may be more informative in future work to enroll participants that reflect a full range of WMCL and evaluate correlations between WMCL and inferencing outcomes.

Another potential issue is the design and duration of the study, which was a single session of approximately one hour and thirty minutes. As observed by the PI during the session and in the transcripts, the participant's responses got shorter in length and in elaborative details as the session progressed. Also, the participants made anecdotal remarks to the PI throughout the session (e.g., in talking about time in one of the experimental scenarios, adding "much shorter than this [the experimental session] is taking," or asking "how much longer will this take"). Thus the results may have been affected by participants' fatigue, boredom, or just wanting to be finished. In the design of the study, this potential problem was weighed against the disadvantages of spreading the testing over two sessions. Gathering data in a single session eliminated outside factors that could influence the participants' responses or the chances of participants dropping out of the study, and was practically more manageable for the PI.

Another potential limitation involved the within study design, meaning all participants got all versions of all the materials. Thus it was possible that their response to one version of a stimulus could be influenced by the prior version. Anecdotally, this appeared to be the case for some participants. The consequences of this concern were managed by counterbalancing the order of stimuli versions within each task, but this design feature may have weakened the chances of detecting differences between WMCL groups or relationships with episodic buffer functioning.

Finally, the episodic buffer stimuli recordings were a bit unnatural because there was a lengthy pause (approximately two seconds) between sentences. Although the inter-sentential pause was equal in coherent and scrambled versions of the stimuli, this would be an important to test in order to see if it affected the findings. Future work with this sort of measure should probably use a more natural presentation pace.

4.6 FUTURE DIRECTIONS

There are many factors that account for the vast variability in cognitive and communicative performance in neurotypical adults. This is evidenced in part by the variability in how the working memory system is conceptualized and measured across experiments.

As seen in this study, there are factors that are difficult to control such as implicit and explicit priorities that could be affecting how the participants respond. Therefore finding a more effective and accurate way to measure how participants prioritize their own goals in relation to allocating central executive resources, could allow for a better representation and interpretation of the processes involved.

Further, this investigation shows promise in examining the episodic buffer's role in multiple inference maintenance or narrowing. However, there is still disagreement on how the functioning of the episodic buffer should be measured (Nobre et. al, 2013). In addition, before any measure can be used with confidence, its reliability and validity would need to be established. With the creation of such, more controlled results from the neurotypical population could be gathered and deductions made about inferencing maintenance or narrowing impairments in individuals with neurological disorders, such as left hemisphere damage, right hemisphere damage, or traumatic brain injury.

Finally, other methods of quantifying inferencing should be considered in future work. The thinking outloud task does not capture inferences that participants might draw, consciously or not, but do not ultimately produce.

5.0 CONCLUSIONS

The current study addressed conflicting conclusions about the relationship between working memory and inference maintenance or narrowing. Results indicated that those differing in WMCL, high versus low, did not differ in their maintenance or narrowing of multiple inferences. This study also addressed a conceptual weakness of the prior investigations of working memory and inferencing (e.g. Blake & Lesniewicz, 2005, Blake, 2009), specifically that working memory was equated with the central executive. The results point to a possibly better predictor of inferencing, a measurement of the functioning of the episodic buffer (Baddeley, 2012), which is another component of working memory that has been overlooked in past investigations.

Further investigation into working memory and inferencing should focus on developing a reliable and valid measure of episodic buffer functioning in neurotypical adults. This type of measure could then be applied to inform scientific study of the maintenance and narrowing of inferences in individuals with neurological disorders, such as left hemisphere damage, right hemisphere damage, or traumatic brain injury.

APPENDIX A

SCREENING FORM

(Updates of neurological, sensory, and substance usage status)

Verbal permission to ask screening questions? Y / N

Neurological Status:

"Since we last had contact, have you had any condition that affected your brain?"

IF YES: "What was it?"

IF NO: "Have you ever had a stroke? Hemorrhage? Brain tumor? Parkinson's disease? Alzheimer's disease? Seizures? Schizophrenia? Manic-Depression? A head injury for which you were hospitalized?

"Since we last had contact, have you had a brain scan, such as a CT or MRI of the brain?"

IF YES: "What was the result?"

Sensory Status:

"Since we last had contact, has your hearing changed?"

IF YES: "How has it changed? How much? Are you using a hearing aid?"

IF NO: "Do other people think your hearing is bad?"

Substance Abuse:

"Do you have a problem with alcohol? Drugs? Do other people think you have a problem?"

APPENDIX B

AUDITORY WORKING MEMORY TASK STIMULUS EXAMPLES

<u>Level 2 Sets Example</u>:
Set 1
You sit on a chair.
Trains can fly.

<u>Level 3 Sets Example</u>:
Set 4
Sugar is sweet.
Florida is next to Ohio.
Horses run in the sky.

<u>Level 4 Sets Example</u>:
Set 7
Twelve equals one dozen.
Bicycles are slower than cars.
A book can play.
Feathers can tickle.

<u>Level 5 Sets Example</u>:
Set 10
Carrots can dance.
Fish swim in water.
You sleep on a bed.
You eat breakfast at night.
People have eyes.

APPENDIX C

INFERENCING TASK STIMULUS EXAMPLES

Experimental version: Rita was frustrated with her job waiting on tables. Customers were rude and the chef was impossibly demanding. Rita was near the end of a long day when a rude man complained that the soup she had just served was cold. Rita became irritate as the customer became louder and nastier. She tried her best to regain her control. She impatiently lifted up the bowl of soup. She took the food back to the kitchen. She sighed realizing her shift was almost over. Rita regained her composure and brought a hot bowl of soup for the customer.

Control version: Susan usually liked her job waiting on tables. But today, the customers were impatient and the chef was moody. Near the end of a long and hard day, a man at one of her tables complained that the vegetables were cold. As she became frustrated he apologized for being so nasty. She realized this one complaint was inconsequential. She carefully lifted up the plate of food. She took it back to the kitchen. She sighed realizing her shift was almost over. Susan brought out a hot plate of vegetables, and the customer left a nice tip when he was finished.

APPENDIX D

EPISODIC BUFFER TASK STIMULUS EXAMPLES

Coherent version: As a candidate for magistrate, Jason announced an information booth that he would open at a local fairground. He believed that this would allow him to share his views. Noticing a lemonade stand, he decided he would take a short break, and get a drink. As he returned, he noted a woman who was at his stand. Greeting her, he saw a book that she was carrying. As he gestured towards the book, he forgot the drink he had in his hand. Spilling the drink on woman's book and dress, he declared his apology. Grudgingly, she accepted his apology, and walked away.

Scrambled version: As a candidate for magistrate, Jason announced an information booth that he would open at a local fairground. As she approached, Jason asked her whether she liked the flower box. They explained he was a new boy in school, named Jason. Around midnight, Jason announced his plan to go to bed. They decided they should stop to see what there was. Then, he stated his wish to view her portfolio. Suddenly, he noticed a chipmunk that ran across the sidewalk. Grudgingly, she accepted his apology, and walked away.

APPENDIX E

STIMULI INSTRUCTIONS

Inferencing Task. I'm going to give you some short stories to listen to. I want you to listen to each one, and talk about ideas that come to your mind as you are listening. Tell me about what you are thinking after each sentence of the story.

To help you with this task, I'll give you two things to focus on. First, I want you to talk about any predictions you have about what will happen in the story. Either what will happen next, or the eventual outcome of the story. As you go along, of course you can change or revise your predictions based on new information in the story.

Along with the predictions, you can also talk about any other ideas that come to your mind related to information in the story.

Some of the stories will seem very logical, and will flow smoothly. Other stories, however, may have unexpected events. If there is anything that seems odd or unusual, let me know as you're talking about the story.

There are no right or wrong answers to any of these. We are just looking for your interpretation of the stories. [e]

Episodic Buffer Task. I'm going to give you a set of short stories to listen to. I want you to listen to each one and then recall the story to me afterwards with as much as precision as possible.

Some of the stories will seem very logical and coherent while others may seem scrambled and incoherent. Remember to try to recall the story to me in the order the story is presented. If there is anything that seems odd or unusual, let me know as you recall it to me afterwards.

[e] Modified from the instructions provided by Blake (2005). For the purpose of this study the word reading was changed to listening.

BIBLIOGRAPHY

Baddeley, A.D., & Hitch, G. (1974). Working Memory. In G.H. Bower (Ed.), *The psychology and motivation* (Vol. 8, 47-89). San Diego, CA: Academic Press.

Baddeley, A., Cocchini, G., Della Sala, S., Logie, R. H., & Spinnler, H. (1999). Working memory and vigilance: evidence from normal aging and Alzheimer's disease. *Brain and Cognition*, *41*(1), 87-108.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.

Baddeley, A.D., & Wilson, B. A. (2002). Prose recall and amnesia: implications for the structure of working memory. *Neuropsychologia, 40*(10), 1737-1743.

Baddeley, A.D., Allen, R.J., & Hitch, G. (2010). Investigating the episodic buffer. *Psychologica Belgica, 50*(3-4), 223-243.

Baddeley, A.D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29

Blake, M.L. & Lesniewicz K. (2005). Contextual bias and predictive inferencing in adults with and without right hemisphere brain damage. *Aphasiology*, 19, 423-434

Blake ML. (2009). Inferencing processes after right hemisphere brain damage: Effects of contextual bias. *Journal of Speech, Language and Hearing Research*, 52, 373-384.

Brady, J. L. (2009) Examining inter-sentential influences on predicted verb subcategorization. Doctoral Dissertation, University of Pittsburgh.

Christoffels, I. K. (2006). Listening while talking: The retention of prose under articulatory suppression in relation to simultaneous interpreting. *European Journal of Cognitive Psychology, 18*, 206-220.

Dunn, L. M., & Dunn, L.M. (2000). Peabody Picture Vocabulary Test III. Circle Pines, MN: American Guidance Service.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, *44*, 145-200.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198

Jeffries, E., Lambon Ralph, M.A., & Baddeley, A.D. (2004). Automatic and controlled processing in sentence recall: the role of long-term and working memory. *Journal of Memory and Language*, *51*(4), 623-643.

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

Henry, L. (2011). *The development of working memory in children*. London, UK: Sage Publications.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47.

Kapikian, A., & Briscoe, J. (2012). Semantic binding, not attentional control, generates coherent global structure in children's narrative memory. *European Journal of Cognitive Psychology*, *24*, 751-764.

Nobre, A. D. P., Rodrigues, J. D. C., Sbicigo, J. B., Piccolo, L. D. R., Zortea, M., Junior, S. D., & de Salles, J. F. (2013). Tasks for assessment of the episodic buffer: A systematic review. *Psychology & Neuroscience*, *6*(3), 331.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Tompkins, C.A., Bloise, C. G. R., Timko, M. L., & Baumgaertner, A. (1994). Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech and Hearing Research*, *37*, 896-912.

Tompkins, C. A., Fassbinder, W., Blake, M. L., Baumgaertner, A., & Jayaram, N. (2004). Inference generation during text comprehension by adults with right hemisphere brain damage: Activation failure vs. multiple activation? *Journal of Speech, Language, and Hearing Research*, 47, 1380-1395.