DO COMPLEX SPAN AND CONTENT-EMBEDDED WORKING MEMORY TASKS
PREDICT UNIQUE VARIANCE IN INDUCTIVE REASONING?

A thesis submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

by
Amanda Sue Zamary

August 2017

Thesis written by

Amanda Sue Zamar

B.S., Otterbein University, 2014

M.A., Kent State University, 2017

Approved by

_____, Advisor
Dr. Katherine Rawson

_____, Chair, Department of Psychological Sciences
Dr. Maria Zaragoza

_____, Dean, College of Arts and Sciences
Dr. James Blank

TABLE OF CONTENTS

LIST OF TABLES	IV
LIST OF FIGURES	V
ACKNOWLEDGEMENTS	VI
CHAPTER 1 INTRODUCTION	1
1.1 What is Working Memory	2
1.2 The Involvement of Working Memory in Inductive Reasoning	5
1.3 Measurement of Working Memory	8
1.4 The Current Research	10
CHAPTER 2 METHOD	12
2.1 Participants	12
2.2 Materials and Procedure	12
CHAPTER 3 RESULTS	19
3.1 Preliminary Analyses	20
3.2 Structural Equation Modeling	21
CHAPTER 4 GENERAL DISCUSSION	25
REFERENCES	29

LIST OF TABLES

Table 1. Means, standard deviations, and correlations of the nine measured variables...	35
--	----

LIST OF FIGURES

Figure 1. Hypothesized model displaying standardized parameter estimates.....	36
--	----

ACKNOWLEDGEMENTS

First and foremost, thank you to Katherine Rawson for supporting my research interests and discussing working memory theory with me on numerous occasions. This thesis would not have been possible without your constructive feedback and mentorship. Also, thank you to Chris Was for providing consultation about working memory and Structural Equation Modeling. Finally, thank you to John Dunlosky and Jeff Ciesla for serving on my thesis committee.

Amanda Zarnary

June 27, Kent, Ohio

CHAPTER 1

Introduction

People use inductive reasoning to make inferences and solve problems on a daily basis. Inductive reasoning involves reasoning from the particular to the general; it is an explicit process that involves the discovery of common relationships among stimuli elements via the formation and testing of hypotheses within a stimulus set (Carroll, 1993; Ekstrom, French, Harman, & Dermen, 1976; Johnson-Laird, 2013; Klauer & Phye, 2008; Klauer, Willmes & Phye, 2002). During inductive reasoning, multiple elements and/or relations between elements are attended to and manipulated to derive a solution (Johnson-Laird, 2013; Klauer & Phye, 2008). Reasoners can adopt various strategies during inductive reasoning, such as systematically comparing stimuli elements and relations between elements, a more heuristic approach in which the problem is examined globally and plausible hypotheses are generated and tested, or iterative combinations of more than one strategy (Klauer & Phye, 2008). Regardless of the strategy used, maintaining task-relevant information during processing is important for inductive reasoning to take place (Cowan, 1988; Johnson-Laird, 2013; Oberauer, 2002; Oberauer, Suß, Wilhelm, & Sander, 2007; Sternberg, 1986; Sternberg & Gardner, 1983).

Inductive reasoning theories assume that maintenance is achieved by the working memory system (e.g., Johnson-Laird, 2013; Sternberg & Gardner, 1983; Sternberg, 1986). However, other facets of the working memory system may also be important for inductive reasoning. Given that the working memory system is multifaceted, different

kinds of working memory tasks may reflect various components of the working memory system to different degrees. The purpose of the current research was to investigate the extent to which different kinds of working memory tasks predict unique variance in inductive reasoning. As a brief overview, we begin by discussing working memory theory and then how working memory is involved in inductive reasoning. Next, we discuss measurement of working memory and predictions concerning the extent to which two kinds of working memory tasks (complex span and content-embedded tasks) will predict unique variance in inductive reasoning.

1.1 What is Working Memory

In a recent review, Cowan (in press) notes that the working memory literature is convoluted given the numerous definitions of working memory adopted by researchers. Cowan discusses some of the common definitions of working memory and recommends that researchers explicitly state the definition they are adopting to improve communication and further research in the field. Following this call for specificity, we adopt what Cowan refers to as the generic working memory definition (Cowan, 1988; Cowan, in press), which states that working memory is “the ensemble of components of the mind that hold a limited amount of information temporarily in a heightened state of availability for use in ongoing information processing” (Cowan, in press, p. 6).

Although explicitly stating a functional definition of working memory is important for communication purposes, Cowan also notes the important distinction between definition and theory. Researchers may adopt the same functional definition of working memory (i.e., what working memory does) but have different theories about the

specific processes and components of working memory that achieve that function. With that said, most theories of working memory assume that working memory is a multifaceted system (see Miyake & Shah, 1999 for perspectives on the non-unitary nature of working memory) and show some consensus in general processes that contribute to the function of working memory. In particular, two broad types of mechanisms that contribute to ongoing cognitive processing include maintenance and disengagement (Shipstead, Harrison, & Engle, 2016). *Maintenance* keeps contents of working memory at a stable, heightened state of activation for ongoing processing (Cowan, 1988; Oberauer et al., 2007; Shipstead et al., 2016). *Disengagement* involves the intentional removal of outdated or irrelevant content from working memory (Oberauer et al., 2007; Shipstead et al., 2016).

To illustrate how processes of maintenance and disengagement are involved in the working memory system, consider the concentric model of working memory (Oberauer, 2002; Oberauer et al., 2007; Wilhelm, Hildebrandt, & Oberauer, 2013). The concentric model theorizes that working memory maintains and processes information in a three-layer system, which includes the region of direct access, the focus of attention, and a highly activated portion of long-term memory. In the region of direct access, stimuli elements and relationships are held at a highly stable, easily accessible state for ongoing processing. Elements held in the region of direct access may include external information from the environment and internal information from long-term memory.

The concentric model assumes that elements held in the region of direct access are organized in a temporary coordinate system that includes a limited number of

placeholders for elements and the connections between them (Oberauer et al., 2007). The nature of this coordinate system varies given current cognitive task demands (e.g., the structure may provide temporal order, spatial organization, or a causal system of elements). The focus of attention selects one element at a time from the region of direct access to manipulate and update (i.e., process). Concerning the highly activated portion of long-term memory, the model assumes that memory representations in long-term memory vary in activation level (in which higher levels of activation leads to a greater likelihood of retrieval into the region of direct access). No limits in the amount of information that can be activated in long-term memory at once exist.

Concerning maintenance, the concentric model of working memory proposes that binding is the mechanism that temporarily connects elements to placeholders in the region of direct access, which assists in keeping information highly stable and accessible for ongoing processing. Maintaining elements in the region of direct access supports construction of new mental representations and integration of information into long-term memory. Elements that are maintained in the region of direct access also serve as cues that raise the activation level of related information in long-term memory. Therefore, as the number of elements and relationships in the region of direct access increases, the amount of potentially relevant information in long-term memory held at higher levels of activation also increases. Having more information held at high levels of activation in long-term memory can assist cognitive performance, to the extent that task-relevant memory representations are more easily retrieved during processing.

Concerning disengagement, intentionally disengaging from currently irrelevant elements in the region of direct access is important for multiple reasons. Most critically, given that the focus of attention selects information from the region of direct access for processing and that the region of direct access is space-limited, disengaging from irrelevant elements and relationships frees space for more task-relevant information (Oberauer et al., 2007). Additionally, disengaging from irrelevant elements will decrease interference from information in activated long-term memory, to the extent that irrelevant elements in the region of direct access served as cues to activate irrelevant information in long-term memory. If too many representations are held at a highly activated state in long-term memory, retrieving task-relevant information may be difficult due to high interference (Oberauer et al., 2007).

The concentric model of working memory is just one of many working memory theories that discuss mechanisms related to maintenance and disengagement. Many other theories view maintenance and disengagement as critical components of the working memory system that contribute to ongoing cognitive processing (see also Cowan 1988; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Wilhelm et al., 2013).

1.2 The Involvement of Working Memory in Inductive Reasoning

To revisit, most theories of inductive reasoning assume that the working memory system is involved in inductive reasoning (e.g., Johnson-Laird, 2013; Sternberg & Gardner, 1983; Sternberg, 1986). Consistent with this assumption, a wealth of research shows strong, positive relationships between inductive reasoning and working memory

(e.g., Ackerman, Beier, & Boyle, 2002, 2005; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Kane, Hambrick & Conway, 2005; Oberauer et al., 2007; Unsworth & Engle, 2005). Importantly, most evidence indicates that reasoning (including inductive reasoning) and working memory are highly related but clearly separable constructs (Ackerman et al., 2005; Kyllonen & Kell, 2017; Oberauer, Schulze, Wilhelm, & Suß, 2005; but see Kyllonen & Christal, 1990). A recent meta-analysis estimated the correlation between reasoning and working memory to be $r = .48$ (Ackerman et al., 2005).

The robust relationship between working memory and inductive reasoning is unsurprising given that maintenance and disengagement functions executed by the working memory system are important for inductive reasoning. To revisit, inductive reasoning involves the discovery of common relationships among stimuli elements via the formation and testing of hypotheses in a series of data. Concerning maintenance, the working memory system maintains problem-relevant stimuli elements, the hypothesis currently being tested, and the relationships between elements that have been identified via hypothesis testing. To illustrate how maintenance supports inductive reasoning, consider the Raven's Advanced Progressive Matrices (Raven, 1962). On each trial, reasoners are given a 3x3 matrix. The first eight cells of the matrix contain figures differing in shape composition, shading, and size and the ninth cell is left empty. The reasoner is given eight additional figures and is asked to identify which figure correctly completes the matrix (as indicated by multiple patterns observed in the set of stimuli). For example, the reasoner may hypothesize that shape size decreases across rows and set

up a coordinate system in the region of direct access aligning with this hypothesis. The reasoner maintains this hypothesis during execution, in which they may compare elements from one cell to another across rows. If the reasoner confirms that shape size does decrease across rows, maintaining this relationship about shape size is important as the reasoner goes back to the problem to detect the other relationships that also determine the correct solution (e.g., the reasoner may next test the hypothesis that shape shading increases down columns). Maintaining all relationships that contribute to the solution is important so that they can be combined to derive a solution (Sternberg, 1986).

Concerning disengagement, intentionally disengaging from problem-irrelevant elements, irrelevant or nonexistent relationships, and incorrect hypotheses is also important for inductive reasoning. To revisit, the region of direct access is limited in space and the focus of attention only selects and processes information in this region. Thus, successfully solving an inductive reasoning problem requires that relevant elements and relationships between elements be brought into the region of direct access for processing. If irrelevant information consumes the region of direct access, inductive reasoning may be impeded because relevant information may not be available for processing in the region of direct access. Additionally, elements in the region of direct access serve as cues to activate information in long-term memory. Thus, irrelevant elements and relationships held in the region of direct access may raise activation levels of irrelevant memory representations in long-term memory. This increase in activation of irrelevant information may interfere with retrieval of memory representations that are relevant for the current problem. Finally, disengaging from unviable hypotheses is

important for successful inductive reasoning. If the reasoner fixates on a hypothesis that failed and subsequently does not test new hypotheses, the likelihood of coming to a correct solution decreases. Shipstead et al. (2016) recently emphasized the importance of disengagement for successful reasoning, given that initial focus on particular stimuli elements may be irrelevant for the problem and initial hypotheses may be wrong. Indeed, some theories even assert that unbinding is the hallmark of reasoning (Oberauer et al., 2007; Shipstead et al., 2016).

1.3 Measurement of Working Memory

According to Shipstead et al. (2016), the associations between working memory and reasoning observed in prior research may have been constrained in part due to the way in which working memory and reasoning are measured. In particular, they note that the tasks most commonly used to measure working memory (i.e., complex span tasks) put a greater emphasis on maintenance processes relative to tasks commonly used to measure reasoning, which put a greater emphasis on disengagement processes.

Complex span tasks are a type of working memory task that involves both storage and processing demands (Cowan, in press; Daneman & Carpenter, 1980). Importantly, the information being stored is different from the information being processed. For example, in the reading span task, participants are presented with sentences one at a time and are asked to identify if the sentence makes sense (i.e., the processing component task). After each sentence, participants are shown a word to remember for later recall (i.e., the storage component task). After each block of sentences, participants are asked to recall the to-be-remembered words in serial order. Although participants are told to

complete both tasks as accurately as possible, working memory is measured as performance on the storage task. Shipstead et al. (2016) note that complex span tasks put heavy emphasis on maintenance processes and place less emphasis on disengagement processes. They state, “the most salient obstacle to accurate task performance is an interpolated processing task, which displaces to-be-remembered information from focal attention. Within the context of this type of task, maintenance is at a premium” (p. 783).

This analysis raises the possibility that other working memory tasks that place a greater emphasis on disengagement processes may predict more variance in inductive reasoning. One type of working memory task that puts a greater emphasis on disengagement processes are content-embedded tasks (Ackerman et al., 2002; Kyllonen & Cristal, 1990; Was, Rawson, Bailey, & Dunlosky, 2011; Woltz, 1988). Similar to complex span tasks, content-embedded tasks also involve storage and processing demands. In contrast to complex span tasks, content maintained in working memory is the same content that is processed. This seemingly minor difference between complex span and content-embedded tasks puts a greater emphasis on disengagement processes in content-embedded tasks.

To illustrate how content-embedded tasks put a greater emphasis on disengagement, consider the ABCD task. On each trial, participants are shown three pieces of information one at a time that specify the ordering of the same four letters (ABCD). The first piece of information states the ordering of the letters A and B (e.g., B comes before A). The second piece of information states the ordering of the letters C and D (e.g., D comes after C). The third piece of information states the ordering of the two

sets of letters (e.g., set 1 comes after set 2). The participant is then asked to indicate the correct solution (CDBA).

Content-embedded tasks involve maintenance of elements and temporary relationships between them to arrive to the correct solution. For example, on the ABCD trial above, participants need to temporarily maintain BA and CD to complete the ordering of the sets as instructed in the third step of the problem. However, disengagement processes are especially important in content-embedded tasks, both within trials and in transitioning from one trial to the next. As an example of how disengagement is important in transitioning between trials, consider again the ABCD task. In this task, the same letters are used on every trial and are presented in the same way (i.e., information about the relationship between A and B, the relationship between C and D, and the relationship between set orders), but the actual relationships between the elements change between trials. Given that the same elements are used on every trial and only relationships change, intentionally disengaging from temporary relationships between elements at the start of each trial is important to reduce interference. Otherwise, lingering relationships from previous trials may make it difficult to maintain relationships in the current trial.

1.4 The Current Research

The purpose of the current research was to investigate the extent to which complex span and content-embedded tasks predict unique variance in inductive reasoning. To revisit, most theories assume that working memory system is multifaceted.

Given that working memory is multifaceted, different working memory tasks may measure facets of the system to different extents. Specifically, complex span tasks put a greater processing emphasis on maintenance, whereas content-embedded tasks put a greater processing emphasis on disengagement.

Given that complex span tasks and content-embedded tasks overlap to some degree in the facets of the system that they measure, we predict that these two kinds of tasks will share some variance that predicts inductive reasoning. However, given that complex span tasks and content embedded tasks measure maintenance and disengagement to different degrees, tasks may uniquely predict inductive reasoning to the extent to which inductive reasoning is reliant on the processes that the tasks emphasize. Specifically, given that disengagement is especially important for reasoning (e.g., Shipstead et al., 2016), we predict that content-embedded tasks will explain a substantial amount of unique variance in inductive reasoning.

CHAPTER 2

Method

2.1 Participants

The current study included 388 students from a large Midwestern university (76% female; 79% white, 16% black, 5% Asian, 4% first nations, 2% Hispanic or Latino, 1% native Hawaiian or Pacific Islander); 41% were in their first year of college ($M = 2.2$, $SE = 0.1$) and 33% were Psychology majors. Mean age of participants was 19.9 years ($SE = 0.1$). Participants were recruited from the Psychology Department's participant pool and received course credit for participation.

2.2 Materials and Procedure

Complex span tasks. The complex span tasks used in the current research were versions of span tasks described in Kane et al. (2004). Scores on all complex span tasks were computed using weighted, partial credit scoring (see Conway et al., 2005 for discussion). Additionally, we used serial recall scoring; participants only received credit for items recalled in correct ordinal positions. Scores on all complex span tasks were entered into the model as percent correct.

Each trial of the reading span task (RSPAN) included a set of sentences. Set size ranged from two to six sentences. Sentences were presented individually and participants were asked to read each sentence silently and then click a button to indicate if the sentence made sense (e.g., "Mr. Owens left the lawnmower in the lemon"). Across all

trials, half of the sentences made sense and half did not. If the participant did not respond within four seconds, the computer automatically moved them forward. After each sentence, participants were presented with an unrelated word (e.g., eagle) for one second that they were asked to remember for later recall. At the end of the sentence set, participants were prompted to recall the words in the order in which they were presented. Participants completed 15 trials, with one trial of each set size in each of 3 blocks. Trials were presented in a fixed random order within each block.

Each trial of the operation span task (OSPAN) included a set of mathematical expressions. Set size ranged from two to five mathematical expressions. Mathematical expressions were presented individually and participants were asked to read each expression silently and then click a button to indicate if it was correct (e.g., “Is $(4 \times 2) + 5 = 10$?”). Across all trials, half of the expressions were correct and half were not. If the participant did not respond within four seconds, the computer automatically moved them forward. After each expression, participants were presented with a word (e.g., phone) for one second that they were asked to remember for later recall. At the end of each set of mathematical expressions, participants were prompted to recall the words in the order in which they were presented. Participants completed 12 trials, with one trial of each set size in each of 3 blocks. Trials were presented in a fixed random order within each block.

Each trial of the counting span task (CSPAN) included a set of arrays. Set size ranged from two to six arrays; each array was presented individually for as much time as the participant needed. However, participants were told to complete each array as quickly as possible. Each array was composed of a random assortment of squares and circles,

including three to nine dark blue circles, a varying number of light blue circles, and a varying number of dark blue squares (arrays were the same across participants). Participants were asked to count the dark blue circles in each array, clicking on each one as they counted. A checkmark appeared on the circle to show the participant that that circle was counted. After they finished counting the dark blue circles in the array, a new array appeared. Participants were asked to remember the number of dark blue circles in each array for later recall. At the end of the array set, participants were prompted to recall the numbers in the order in which they were presented. Participants completed 15 trials, with one trial of each set size in each of 3 blocks. Trials were presented in a fixed random order within each block.

Content-embedded tasks. The content-embedded tasks used in the current research were versions of content-embedded tasks that have been used as measures of working memory in previous research (e.g., Ackerman et al., 2002; Kyllonen & Christal, 1990; Was et al., 2011; Was & Woltz, 2007; Woltz, 1988). Scores on all content-embedded tasks were computed as number of correct responses per minute, given that prior research using content-embedded tasks suggest that meaningful individual differences are contained in both speed and accuracy (see Vandierendonck, 2017; Was & Woltz, 2007; Woltz & Was, 2006; 2007).

On each trial of the ABCD task, participants were required to process three pieces of information to determine the ordering of four letters (A, B, C, and D). First, participants were given the ordering of letters A and B (e.g., B comes before A). Participants clicked a button to replace the first statement with one giving the ordering of

letters C and D (e.g., D comes after C). Participants again clicked a button to replace the second statement with one giving the ordering of the two pairs of letters (e.g., Set 1 comes after Set 2). Participants clicked a button to advance to the next screen, which showed the eight possible orderings of A, B, C, and D. Participants were asked to select the correct answer and the cycle repeated for the next trial. All screens on each trial were self-paced; however participants were told to respond as quickly as possible. Participants completed 23 trials; letter and set ordering varied by trial and were presented in the same fixed, random order across participants.

On each trial of the Alphabet task, participants were asked to transform sets of letters. Participants were presented with one or two non-adjacent letters from the alphabet with a transformation direction and number (e.g., T forward 3; OZ backward 2; the answers are W and MX, respectively). Once participants solved the transformation, they clicked a button to advance to the next screen, which included eight response options. Participants had up to five seconds to select the correct answer; if they did not select an answer they were automatically moved forward and the trial was counted as incorrect. Participants completed 12 trials in each of two blocks (each block contained one and two-letter trials). Letters and transformations varied by trial and were presented in the same fixed, random order across participants.

On each trial of the Digit task, participants were asked to answer one or two questions about a string of numbers. Participants were presented with six single digit numbers for two seconds each (e.g., 5, 8, 1, 4, 9, 8). After the presentation of the digit string, participants were asked one or two questions about the number string (e.g., “How

many even numbers were there?”, “What is the smaller of the middle two numbers?”). If the trial involved two questions, questions were presented individually. All answers were numeric and participants answered by typing in the correct answer. This phase of the task was self-paced, but participants were asked to answer as quickly and accurately as possible. Participants completed a block of 12 single question trials and then a block of 12 double question trials). Questions varied by trial and were presented in a fixed, random order across participants.

Inductive reasoning tasks. Scores on all three inductive reasoning tasks were computed as percent correct. We used the short form of the Raven’s Advanced Progressive Matrices (RAPM; Raven, 1962, Set II) used by Stanovich and Cunningham (1992; in brief, they dropped 18 of the easiest and most difficult items, given frequent floor and ceiling effects in college students). On each trial, participants saw a 3x3 matrix with the first eight cells containing figures differing in shape composition, shading, and size. Eight additional figures were presented below the matrix. Participants were asked to click on the figure that correctly completed the pattern in the matrix. Participants could complete up to 18 trials and were given up to 12 minutes to complete the task. Trials were presented in ascending order from least to most difficult.

On each trial of the Locations task (Carroll, 1993; Ekstrom et al., 1976), participants were asked to extract a pattern from an array of Xs and dashes. Each array included four rows and each row contained sets of dashes with an X inserted within one of the sets (e.g., - - - - - - - X - - - - - - - - - -). The placement of the X in each row was determined by an unstated rule (e.g., X replaces the third dash in the second set of

each row). Below the array, participants were presented with a fifth line that included sets of dashes with the numbers 1 through 5 dispersed in five locations. Participants were asked to figure out the rule and then select the number that indicated where the X should be placed given the rule. Participants were instructed that the task goal was to get a high score on the test, but to skip a problem if they were unsure of the answer because they would be penalized for answers that were incorrect. Participants could complete up to 14 trials in each of two blocks and had up to five minutes to spend on each block of trials. If a participant skipped one or more trials and had time left within the five minute block, the trial was re-presented until either the participant selected an answer or five minutes were up.

In each trial of the Letter Sets task (Carroll, 1993; Ekstrom et al., 1976), participants received five sets of four letters (e.g., NOPQ, DEFL, ABCD, HIJK, UVWX). A rule determined the composition of four of the sets of letters and one set did not follow the rule (e.g., DEFL does not follow the rule that letters be in alphabetical order). Participants were asked to figure out the rule and then click on the set of letters that did not follow the rule. Participants were instructed that the task goal was to get a high score on the test, but to skip a problem if they were unsure of their answer because they would be penalized for answers that were incorrect. Participants could complete up to 15 trials in each of two blocks and had up to five minutes to spend on each block of trials. If a participant skipped one or more trials and had time left within the five minute block, the trial was re-presented until either the participant selected an answer or five minutes were up.

Data reported were collected as a part of a larger individual differences study. Participants completed additional tasks that are not relevant for present purposes and will not be reported in this manuscript.¹ The entire study involved four sessions across a two week period. Participants did not complete more than one task for any given latent factor during the same session (Session 1: Alphabet and Locations; Session 2: ABCD, OSPAN, and RAPM; Session 3: RSPAN and Letter Sets; Session 4: Digit and CSPAN).

¹ For purposes of full disclosure (see Simmons, Nelson, & Simonsohn, 2011), other tasks that were collected included example-based concept learning tasks and final comprehension tests, as well as individual differences measures of verbal ability and spelling. These data have not yet been analyzed and will be reported elsewhere.

CHAPTER 3

Results

Prior to conducting analyses, data were examined for attrition and evidence of non-compliance. Participants were excluded from analysis if they had more than one missing value from a single latent factor due to attrition ($n = 36$) or due to computer error ($n = 1$). Of the remaining 351 participants, 13 participants were excluded from analysis given evidence of non-compliance on more than one measured variable on a single latent factor [i.e., RSPAN and OSPAN: did not respond to more than 60% of the processing trials; Alphabet, Locations, and Letter Sets: spent less than 90 seconds on the entire task (including instructions); RPM: spent less than 120 seconds on the entire task (including instructions and practice problems)]. Instead of excluding participants who showed non-compliance on a single measure, we treated that single measured variable as missing data. Thirty-four participants were missing data but were included in analysis given that they were missing no more than one measured variable per latent factor ($n = 14$ due to attrition, $n = 19$ due to non-compliance, $n = 1$ due to a lost data file). No more than 4% of data were missing for each measured variable. Less than 2% of data were missing from the entire set of data.

The final sample included 338 participants. Given that we used structural equation modeling and parameter estimates were derived using maximum likelihood, a minimum

of five cases per parameter estimate is recommended (Mueller & Hancock, 2010). Our sample size well exceeds the minimum requirement for the model to be tested (i.e., 21 estimated parameters, with 16 cases per parameter). All analyses were conducted in MPlus version 7.31 (Muthén & Muthén, 2015).² Values for missing data were estimated using full information maximum likelihood.

3.1 Preliminary Analyses

Table 1 includes summary statistics, zero order correlations, and reliability estimates for measured variables. Importantly, the three measures composing each latent factor correlated highly with each other.³ We also screened for univariate normality and multivariate normality. Concerning univariate normality, skewness statistics on each measured variable were all smaller than 1.5 and kurtosis statistics were all smaller than 3.1, meeting the assumption of univariate normality for the use of maximum likelihood. Concerning multivariate normality, Mardia's measures of multivariate skewness and

² Primary analyses were also conducted in AMOS version 22 (Arbuckle, 2013). Values did not deviate from results found in MPlus.

³ The one exception is the locations task. Zero-order correlations between locations and the other inductive reasoning tasks were somewhat weaker than expected. To foreshadow, the task significantly loads onto the inductive reasoning factor (although the loading was numerically weaker than expected). All three of these inductive reasoning tasks have been used to compose a single-latent factor in previous research, in which the locations task loaded more strongly (Was, Dunlosky, Bailey, & Rawson, 2012). Most importantly, neither of the working memory factors would be differentially disadvantaged by this loading, given that the task is a part of the inductive reasoning factor.

kurtosis were significant ($z = 648.74$, $p < .001$ and $z = 10.19$, $p < .001$, respectively), indicating multivariate non-normality.⁴

3.2 Structural Equation Modeling

Both complex span tasks and content-embedded tasks were expected to predict inductive reasoning; accordingly, the model included paths for both of these directional effects. Additionally, given that both kinds of working memory tasks are predicted to measure some of the same facets of the working memory system, complex span tasks and content-embedded tasks were expected to correlate with one another. Accordingly, the model included a path for this non-directional effect. The hypothesized model with standardized path coefficients and estimated factor correlations is presented in Figure 1.

Concerning model fit, the chi-square test of model fit indicated that the model did not fit the data well [$\chi^2(24) = 46.16$, $p = .004$]. However, two limitations of the chi-square test of model fit include (1) that it assumes multivariate normality and even slight deviations from the specified model may produce large chi square values and (2) that it is overly strict when sample size is large (Bentler & Bonnet, 1980; McIntosh, 2006). Given that the multivariate normality assumption was not met and that sample size is large, other model fit indices are more appropriate. Importantly, all other model fit indices

⁴ To ensure that multivariate non-normality did not affect the qualitative pattern of findings, we also calculated estimates for all models using 500 bootstrap samples. Parameter estimates were similar following bootstrapping and 95% confidence intervals for the standard error of the regression coefficients indicated that significant parameter estimates were not affected by the bootstrap sampling.

indicated that the model fit the data well [CFI: .97, SRMR: .04, RMSEA: .05, $p = .41$, 90% CI: (.03, .08)]. All measured variables significantly loaded onto their respective latent factor and latent factors were strongly correlated with one another.

All model relationship statistics are reported using standardized estimates. As predicted, complex span and content-embedded latent factors were strongly correlated ($r = .74$, $p < .001$). Of primary interest, we predicted that content-embedded tasks would strongly predict inductive reasoning, given that content-embedded tasks put emphasis on disengagement processes. Indeed, content-embedded tasks uniquely predicted inductive reasoning [$\beta = .67$, $SE = .13$, $p < .001$; 95% CI: (.43, .92)]. Interestingly, complex-span tasks did not uniquely predict inductive reasoning [$\beta = .06$, $SE = .13$, $p = .65$; 95% CI: (-.19, .31)]. In total, the model predicted 51% of the variance in inductive reasoning: 45% of the variance was uniquely explained by content-embedded tasks, 6% of the variance was due to overlapping variance between the working memory task types, and less than 1% of the variance (.004%) was uniquely explained by complex span tasks.

We also investigated whether the relationship between complex span tasks and inductive reasoning was significantly attenuated when content-embedded tasks were included in the model. For this model, we constrained the path between complex span tasks and inductive reasoning to the parameter estimate when the content-embedded factor was not included in the model. When the model only included inductive reasoning regressed on complex span tasks, complex span tasks significantly predicted inductive reasoning ($b = .23$, $SE = .05$, $p < .001$). A chi square difference test between the freely

estimated, hypothesized model and the fixed parameter model indicated that model fit of the unconstrained model was better than model fit of the constrained model [$\Delta\chi^2(1) = 11$, $p < .01$].⁵ Additionally, other model fit indices indicated poorer or same fit in the constrained model relative to the unconstrained model [CFI: .96, SRMR: .04, RMSEA: .06, $p = .17$, 90% CI: (.04, .08)].

These findings suggest that reliable change in the relationship between complex span tasks and inductive reasoning was observed when content-embedded tasks were included in the model. When complex span tasks were the only working memory task type in the model, complex span tasks significantly predicted inductive reasoning. However, when content-embedded tasks were also included in the model, complex span tasks no longer uniquely predicted inductive reasoning. Instead, the vast majority of variance that was predicted by complex span tasks was represented by overlapping variance between content-embedded and complex span tasks.

We also compared our hypothesized model to a model in which all working memory tasks were loaded onto a single working memory factor to ensure that this more parsimonious model did not fit the data better. Model fit statistics indicated worse fit in this single factor model compared to our hypothesized model [$\chi^2(26) = 105.98$, $p < .001$, CFI: .90, SRMR: .05, RMSEA: .10, $p < .001$, 90% CI: (.08, .12)]. A chi square difference

⁵ We also tested for attenuation using the Sobel test. In our model, this test produced a z-statistic of $z = 7.66$, $p < .001$, also suggesting that the effect of complex span tasks on inductive reasoning is attenuated when content-embedded tasks are also in the model.

test comparing the two models suggested that the model is oversimplified when working memory tasks are loaded onto a single factor [$\Delta\chi^2(2) = 59.81, p < .01$]. Additionally, AIC values were smaller in hypothesized model versus the single factor model (19928 versus 19984) and less variance in inductive reasoning was explained in the single factor model (47%), all suggesting that the hypothesized model should be retained.

CHAPTER 4

General Discussion

The current research was the first study to examine the extent to which content-embedded tasks and complex span tasks predict unique variance in inductive reasoning. To revisit, these types of tasks differentially measure aspects of the working memory system. Complex span tasks put a greater processing emphasis on maintenance processes (responsible for keeping information at a highly activated, stable state for processing) whereas content-embedded tasks put a greater emphasis on disengagement processes (responsible for the intentional removal of outdated or irrelevant information in working memory). Given that disengagement is particularly important for inductive reasoning, we predicted that content-embedded tasks would explain a substantial amount of unique variance in inductive reasoning. Confirming our prediction, the model explained 51% of the variance in inductive reasoning and 45% of the total variance was uniquely explained by content-embedded tasks. Less than 1% was uniquely explained by complex span tasks.

Although theories of inductive reasoning generally assume that the working memory system supports inductive reasoning processes, they have focused less on how the working memory system assists in reasoning (Klauer & Phye, 2008; Klauer, Willmes & Phye, 2002; Sternberg, 1986; Sternberg & Gardner, 1983). Of the inductive reasoning

theories that discuss working memory, most discuss the value of working memory in reasoning as maintaining relevant information (e.g., Pellegrino & Glaser, 1982; Sternberg & Gardner, 1983). These theories less often discuss how disengagement assists in inductive reasoning [but see Holyoak, 2013 for discussion concerning disengagement in analogical reasoning (a subtype of inductive reasoning)]. This point is interesting, given that the working memory literature asserts that disengagement processes are perhaps more important than maintenance processes for reasoning (Oberauer et al., 2007; Shipstead et al., 2016). To the extent that the unique variance predicted by content-embedded tasks reflects measurement of disengagement processes, the findings from the current research have important implications for theories of inductive reasoning. Specifically, findings from the current research suggest that disengagement processes (executed by the working memory system) may be particularly important for inductive reasoning.

The reasoning tasks in the current research were inductive reasoning tasks; however, findings may generalize to reasoning more broadly. A unified theory of reasoning suggests that inductive and deductive reasoning rely on the same underlying processes (Sternberg, 1986) and working memory theories have discussed the importance of disengagement for reasoning more broadly (Oberauer et al., 2007; Wilhelm et al., 2013) and even fluid intelligence (Shipstead et al., 2016). Although reasoning performance may be better predicted by content-embedded tasks compared to complex span tasks, performance on other cognitive tasks that are more reliant on maintenance

processes may be better predicted by complex span tasks. The broader implication suggested by the current research is that for any given task domain, carefully considering the processes in the working memory system that will be predictive of the outcome of interest is important. After considering which processes will be particularly important, researchers can select working memory tasks that emphasize facets of the working memory system motivated by their theory. Motivating task selection by theory will allow for the researchers to have a better measure of working memory in the context of their research domain, resulting in a better working memory predictor.

Complex outcome measures may rely more heavily on multiple components of the working memory system. In this case, using multiple types of working memory tasks that measure different facets of the working memory system may be beneficial. If researchers use multiple types of working memory tasks that emphasize different facets of the working memory system, they can extract latent factors associated with theorized facets and investigate the influence of specific facets and the working memory system as a whole on their outcome of interest.

Perhaps the most surprising finding in the current research was that complex span tasks predicted virtually no unique variance in inductive reasoning. Although we expected that content-embedded tasks would predict substantial unique variance in inductive reasoning, we did not expect that complex span tasks would not also predict unique variance in inductive reasoning. This finding was surprising given that complex span tasks put a heavily processing emphasis on maintenance (Shipstead et al., 2016),

which also supports inductive reasoning. However, maintenance processes may have been captured by overlapping variance between content-embedded tasks and complex span tasks given that content-embedded tasks also reflect maintenance to some degree. If this speculation is true, our findings are still consistent with the idea that inductive reasoning is more reliant on disengagement processes than maintenance processes. Only 6% of the explained variance in inductive reasoning was due to shared variance between content-embedded tasks and complex span tasks, whereas 45% of the explained variance was uniquely due to content-embedded tasks.

In sum, the current research shows that content-embedded tasks predict a substantial amount of unique variance in inductive reasoning, whereas complex span tasks predicted almost no unique variance in inductive reasoning. We attribute the unique predictive power of content-embedded tasks to greater emphasis on disengagement processes, which play an important role in successful inductive reasoning.

REFERENCES

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131(1), 30 – 60.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131, 567 – 589.
- Arbuckle, J. L. (2013). AMOS (Version 22) [Computer software]. Chicago: IBM SPSS.
- Bentler, P. M. & Bonnet, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588 – 606.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163 – 183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769 – 786.

- Cowan, N. (in press). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*. DOI: 10.3758/s13423-016-1191-6.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163 – 191.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- Ecker, U. K. H., Lewandowsky, S., Oberauer, K., Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 170 – 189.
- Ekstrom, R. B., French, J. W., & Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- Holyoak, K. J. (2013). Analogy and relational reasoning. In Holyoak, K. J. & Morrison, R. G. (Eds.), *The oxford handbook of thinking and reasoning* (pp. 234 – 259). New York, NY: Oxford University Press.
- Johnson-Laird, P. N. (2013). Inference in mental models. In Holyoak, K. J. & Morrison, R. G. (Eds.), *The oxford handbook of thinking and reasoning* (pp. 134 – 154). New York, NY: Oxford University Press.
- Kane, M. J., Hamrick, D. Z., Tuholsky, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable

- approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189 - 217.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on ackerman, beier, & boyle (2005). *Psychological Bulletin*, 131(1), 66 - 71.
- Klauer, K. J. & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research*, 78(1), 85-123.
- Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, 27, 1 – 25.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, 14(4), 389-433.
- Kyllonen, P. & Kell, H. (2017). What is fluid intelligence? Can it be improved? In Rosen, M., Hansen, K. Y., & Wolff, U. (Eds.), *Cognitive abilities and educational outcomes* (pp. 15 – 38). Switzerland: Springer International Publishing.
- McIntosh, C. (2006). Rethinking fit assessment in structural equation modeling: A commentary and elaboration on barrett (2007). *Personality and individual differences*, 42(5), 859 – 867.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49 – 100.

- Miyake, A. & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press.
- Mueller, R. O. & Hancock, G. R. (2010). Structural Equation Modeling. In G. R. Hancock & R. O. Mueller (Eds.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (pp. 371 – 384). New York, NY: Routledge.
- Muthén, L. K., & Muthén, B. O. (2015). MPlus (Version 7.31) [Computer software]. Los Angeles, CA.
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 411 – 421.
- Oberauer, K., Schulze, R., Wilhelm, O., & Suß, H. M. (2005). Working memory and intelligence – their correlation and relation: comment on ackerman, beier, and boyle (2005). *Psychological Bulletin*, 131(1), 61 – 65.
- Oberauer, K., Suß, H., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York, NY: Oxford University Press.
- Pellegrino, J. W. & Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning. In R. Glaser (Ed.), *Advances in instructional psychology*. Hillsdale, NJ: Erlbaum.
- Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis.

- Shipstead, Z., Harrison, T. L., & Engle, R. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771 – 799.
- Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359 – 1366.
- Stanovich, K. E. & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition, 20*(1), 51 – 68.
- Sternberg, R. J. (1986). Toward a unified theory of human reasoning. *Intelligence, 10*, 281 – 314.
- Sternberg, R. J. & Gardner, M. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*(1), 80 – 116.
- Unsworth, N. & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence, 33*, 67 – 81.
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods, 49*(2), 653 – 673.
- Was, C. A., Dunlosky, J., Bailey, H., & Rawson, K. A. (2012). The unique contributions of the facilitation of procedural memory and working memory to individual differences in intelligence. *Acta Psychologica, 139*, 425 – 433.

- Was, C. A., Rawson, K. A., Bailey, H., & Dunlosky, J. (2011). Content-embedded tasks beat complex span for predicting comprehension. *Behavior Research Methods*, 43(4), 910-915.
- Was, C. A. & Woltz, D. J. (2007). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language*, 56, 86 – 102.
- Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it?. *Frontiers in Psychology*, 4, 433.
- Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, 117, 319 – 331.
- Woltz, D. J. & Was, C. A. (2007). Available but unattended conceptual information in working memory: Temporarily active semantic content or persistent memory for prior operations? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 155 – 168.
- Woltz, D. J. & Was, C. A. (2006). Availability of long-term memory during and after attention focus in working memory. *Memory and Cognition*, 34, 668 – 684.

Table 1

Means, standard deviations, and correlations of the nine measured variables

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9
1. RSPAN	64	19	(.86)	.76	.53	.46	.35	.54	.30	.23	.39
2. OSPAN	81	16	.63	(.79)	.57	.50	.45	.52	.29	.27	.34
3. CSPAN	83	16	.46	.47	(.87)	.42	.25	.47	.37	.28	.34
4. ABCD	3	1	.40	.42	.37	(.89)	.53	.59	.40	.34	.47
5. Alphabet	4	1	.29	.36	.21	.45	(.81)	.52	.18	.25	.44
6. Digit	8	3	.47	.43	.41	.52	.44	(.87)	.30	.31	.41
7. RAPM	34	17	.24	.22	.29	.32	.14	.24	(.72)	.24	.54
8. Locations	38	14	.17	.19	.21	.26	.18	.23	.16	(.64)	.38
9. Letter Sets	55	15	.34	.28	.30	.41	.37	.36	.43	.28	(.87)

Note: Complex span (variables 1- 3) and inductive reasoning scores (variables 7-9) are out of 100%. Content-embedded scores (variables 4- 6) are number correct per minute. All *ps* are < .008. Internal reliability estimates were computed using Cronbach's alpha and are presented on the diagonal (bolded and in parentheses). Observed correlations are presented below the diagonal. Correlations corrected for attenuation are presented above the diagonal.

Figure 1. Hypothesized model displaying standardized parameter estimates. Note: Standardized path coefficients are shown in bold type and estimated factor correlations are shown in parentheses (error variances are not included).