Spring 2019

Evaluating the Latent Variable Structure of Episodic Long-Term Memory Abilities

Kyle Featherston
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological and Brain Sciences

Evaluating the Latent Variable Structure of Episodic Long-Term Memory Abilities

by

Kyle Gramer Featherston

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2019
St. Louis, Missouri

Table of Contents

List of Figures ... iii
List of Tables ... iv
Acknowledgments ... v
Abstract of the Thesis ... vi
Chapter 1: Thesis ... 1
1.1 Evaluating the Latent Variable Structure of Episodic Long-Term Memory Abilities ... 1
1.1.1 Different Kinds of Tasks ... 2
1.1.2 Different Kinds of Items ... 4
1.1.3 Previous Latent Variable Analyses of Episodic LTM ... 5
1.1.4 Current Study ... 7
1.2 Method ... 8
1.2.1 Participants ... 8
1.2.2 Materials ... 11
1.2.3 Procedure ... 13
1.3 Results ... 14
1.3.1 Evaluation of Task Performance ... 15
1.3.2 Exploratory Factor Analysis ... 17
1.4 Discussion ... 23
1.4.2 Conclusion ... 29
References ... 31

List of Figures

Figure 1: Sample Stimuli Presented in Context……………………………………………….11
Figure 2: Scree Plots with Parallel Analysis ………………………..……………...…………19
Figure 3: MAP Index as a Function of Factors ………………………..……………...………19
Figure 4: Path Diagram of the Four Factor Solution………………..……………...…………23

List of Tables

Table 1: Demographic Characteristics…………………………………………………………...10
Table 2: Descriptive Statistics and Powers found for Box-cox Transformations for all Memory Measures………………………..……………...……………………………………...15
Table 3: Correlations between all Memory Measures………………………..……………….18
Table 4: Factor Loadings From Structure Matrix and Communalities for Two Factor Solution..21
Table 5: Factor Loadings From Structure Matrix for Four Factor Solution..………...…………22
Table 6: Model Fit Statistics for Two Factor and Four Factor Solutions……...……...…………22

Acknowledgments

Thank you to my advisor, Sandra Hale, for her guidance throughout the process. Thank you to Joel Myerson for serving on my committee and for his thorough feedback on this paper and to Mike Strube for serving on my committee and providing invaluable statistical instruction and advice. Additional thanks to my Research Assistants Dante Chao, Alyssa Weinrich, and Isaac Kaufer for helping with the initial collection of pilot data for this project.

Kyle Featherston

Washington University in St. Louis

May 2019

ABSTRACT OF THE THESIS

Evaluating the Latent Variable Structure of Episodic Long-Term Memory Abilities

by

Kyle Gramer Featherston

Master of Arts in Psychological and Brain Sciences

Washington University in St. Louis, 2019

Professor Sandra Hale, Chair

I investigated how recall and recognition differ depending on the nature of the memory items and what one is asked to remember about them. Participants were asked to remember lists of various types of verbal items, including words, nonwords, common first names, and the names of common objects in pictures that they viewed, or to remember the contextual information that accompanied those items, including their size, location, color, or font. Immediately following presentation of each list, free recall or recognition tests for items or context were administered. It has been proposed that memory for context, or source memory, differs from episodic memory for items themselves. Exploratory factor analysis suggested that the tasks studied consisted ofthat item recognition and item recall are separate abilities, but did not provide evidence for a separate memory for context.

Chapter 1: Thesis

1.1 Evaluating the Latent Variable Structure of Episodic Long-Term Memory Abilities

Humans can remember numerous things ranging from their home address as a child, to visual memories of what that childhood home looked like, to what phrase started this sentence, to an equation learned in Introductory Statistics. Memory can come in many different forms, based both on what type of information is being remembered and the amount of detail or specifics about the to-be-remembered information that are retrieved. Researchers have proposed many types of memory systems and abilities to attempt to reflect this range of memory types, yet there is still debate about how many unique memory abilities exist.

William James (1890) first proposed that primary memory (i.e., knowledge of information currently held in consciousness) and secondary memory (i.e., knowledge of previous events or facts) differ. Although the terms short-term memory (STM) and long-term memory (LTM) are now favored over primary and secondary memory, the distinction has been supported by research and is largely accepted (see Cowan, 2008, for a review). Moreover, many different memory systems within LTM have been proposed. For example, Tulving (1972) proposed a distinction within LTM between semantic memory, which is knowledge-type memory, and episodic memory, which is memory for events that unfold over time, and this distinction has generally been accepted due to studies showing their dissociation in both young adults (e.g., Nyberg 1994) and older adults (e.g., Mitchell, 1989).

Although episodic LTM is broadly defined as memory for past events, the possible distinctions do not end there. Retrieval of information from the LTM store may differ depending on the level or nature of the details to be remembered (e.g., what was said vs. who said it) as well as the nature of the information that was encoded itself, (e.g., whether it is primarily a visual memory or a verbal one). To a degree this is apparent in the memory literature, as researchers use different tests such as recall, cued-recall, and recognition and have investigated memory for visual, verbal, and auditory information.

There has been surprisingly little research, however, evaluating whether different measures of LTM that use different materials or tests measure different underlying abilities. For example, when using LTM as a predictor or covariate, researchers often give a single or a small number of measures of LTM, assuming that whatever measure is used will do a fine job of approximating the general LTM ability of participants. However, given the lack of consensus in the literature about different memory systems and abilities, this assumption is open to question. One of the goals of experimental psychologists is to take a complex concept and simplify it in the lab, and a simple word list recall task is straightforward and easy to administer, and various aspects of the administration or the stimuli can be manipulated in order to answer research questions. In contrast, psychologists studying individual differences have different goals and take a different approach; one that focuses on accurately measuring abilities in individuals. If there is a single, general episodic LTM ability, then task selection is relatively unimportant, and the reliability of the measure a task provides should be all that matters. Given the diversity of possible memories discussed, however, it is worth further exploring whether episodic LTM actually encompasses multiple, distinct abilities. Therefore, the current study focused on episodic LTM and used diverse, carefully selected tasks in order to evaluate whether some of these possible distinctions are revealed as different latent abilities.

1.1.1 Different Kinds of Tasks

Generally, a measure of episodic LTM involves a study component in which a list of items is presented, and a test component where participants are asked to remember some aspect(s) of the initial list items. From there, many differences emerge between measures. For example, the memory processes that are to be tested may differ. There are generally three different kinds of tests that a researcher might use to assess LTM: recognition, recall, and cued recall. Recognition involves being shown an item and making a judgment about whether it was or was not on the list. Recall involves reporting what can be remembered from the list. Cued recall generally involves being shown one of a set of two items that were paired at study and asking the participant to state/type the other.

Many memory theories distinguish between recall and recognition, and most theorists agree that they are distinct constructs (Gillund & Shriffrin, 1984). One dual-process theory of recognition makes a further distinction and posits that recognition involves a recollection and a familiarity component, whereas recall only involves recollection (Jacoby, 1991; Yonelinas, 2001). There are also dual-process theories of recall, such as the generate-recognize model (Kintsch, 1970), which proposes that recall involves generating potential items from memory and then deciding whether generated items were previously presented, whereas recognition relies solely on this second, decision process. Regardless of exactly which processes are involved in recall and recognition tasks, however, the generally accepted distinction between these two kinds of tasks suggests that a complete study of episodic LTM abilities should include both.

In the majority of studies of episodic LTM, what participants are asked to recall or recognize is the exact item itself. However, as Tulving and Thomson (1973) demonstrated, the context in which an item is encoded is an important factor in memory. Researchers have proposed the construct of source memory (Johnson, Hashtroudi, & Lindsay, 1993), which is measured by testing memory for some aspect of the context in which the to-be-remembered item was initially presented. For example, a spoken word or sentence can be presented and participants tested for recognition of the same or different voice (e.g., Glisky & Polster, 1995; Hicks & Marsh, 1999). Other aspects of context whose recall or recognition can be measured include font (e.g,, Kausler & Puckett, 1980), color (e.g., Doerkson & Shimamur, 2001), location (e.g., Cansino, Maquet, Dolan, & Rugg, 2002), or source list, that is, whether an item was in list A or list B when the lists were encoded at different times or in different ways (e.g., Dobbins, Foley, Schacter, & Wagner, 2002).

The distinction between item and source memory has been supported by the finding of greater age differences in performance for memory for context than items (see Zacks, Hasher, & Li, 2000, for a review) and by neuroimaging studies suggesting that different brain regions are involved (e.g., Slotnick, Moo, Segal, & Hart, 2003; see Mitchell & Johnson, 2009, for a review). However, a study by Siedlecki, Salthouse, and Berish (2005) using confirmatory factor analysis (CFA) to examine age and individual differences suggested that there were not separate latent abilities for memory for content and context, although differences between recall and recognition were not examined. Given the lack of consensus on the content/context distinction, it is worth exploring further whether source memory is, in fact, a separate ability, and whether recall and recognition are associated with differences in that regard.

1.1.2 Different Kinds of Items

The kind of to-be-remembered items most commonly studied are word lists, as they are familiar and easy to present, and aspects of the stimuli such as familiarity and length can easily be controlled. The nature of stimuli that humans need to remember, however, is much more diverse than words and some have suggested that different items may involve different memory abilities. Several researchers have suggested that verbal and visual memory are distinct constructs (e.g., Pavio, 1971), and some research has supported this hypothesis. Verbal and visuospatial working memory tasks load on different factors (Hale et al., 2011), and the results of the studies using a latent construct approach to study episodic LTM memory have provided evidence for separate visual and verbal memory abilities (Hermann et al., 2001; Siedlecki, 2007; Wechsler, 1997).

In addition to using visual and verbal stimuli, verbal stimuli can vary in content. Although episodic memory has been distinguished from semantic memory, the use of lists of words (or pictures of common objects that must be recalled verbally) are obviously associated with meanings that people have knowledge about and thus may rely on in order to help them perform the task (e.g., capitalizing on semantic associations among various items). Given the argument that semantic memory differs from episodic memory, it is a bit odd that episodic memory is most commonly measured with stimuli that have such rich semantics. If what researchers are truly interested in is memory for a specific instance of an event, then it is possible that relying on stimuli for which people have already acquired meaning affects performance.

The familiarity of items to participants can also affect memory performance. For example, Hulme, Maughan, and Brown (1991) demonstrated that simple spans were different depending on the familiarity of the verbal stimuli used. If memory is different on average for familiar and unfamiliar items, there is also the possibility that individuals may rely on different abilities to remember familiar versus unfamiliar items. Unfortunately, there is scarce previous research evaluating difference in LTM abilities for items varying in familiarity.

1.1.3 Previous Latent Variable Analyses of Episodic LTM

Since it was first used in psychological research by Spearman (1927), factor analysis and similar statistical procedures have proven to be useful tools for understanding the underlying abilities affecting performance on different measures. Despite the diverse ways episodic LTM can be measured, surprisingly few studies have systematically examined memory using a latent variable framework. As Hermann et al. (2001) note, most studies of memory have been experimental studies primarily concerned with the validity of proposed distinctions of memory systems (e.g., short-term vs. long-term or episodic vs. semantic). In their analysis of thirteen short-term and LTM tasks, they used Confirmatory Factor Analysis (CFA) to compare eight different models of memory. The best fitting model was a model that separated STM and LTM, and had a distinction of visual, verbal, and semantic memory within LTM. Importantly, there were no models tested that further separated episodic LTM beyond visual and verbal, as well as relatively few tasks that were designed to test episodic LTM, as it was not their primary focus. Siedlecki (2007) used CFA specifically to look at episodic LTM across age ranges and found evidence for different latent LTM abilities depending on the material-type, with a model containing separate factors for verbal, figural, and spatial memory being the best fit. Interestingly, this material-specific model outperformed a model which separated recognition from recall (and cued recall), and only the verbal materials showed evidence of a recall-recognition distinction.

Although relatively few studies have focused specifically on the structure of episodic LTM abilities revealed by latent variables, several studies have used LTM measures as part of broader studies of cognitive abilities. Episodic LTM is a necessary component of many of the complex decisions that human beings need to make, and individual differences in the ability have the potential to predict individual differences in reasoning and problem solving such as those measured by fluid intelligence tasks. Several analyses of the relations between memory measures and intelligences have been done (e.g, Carroll, 1993; Mitchell, 1989; Underwood, Boruch, & Malmi, 1978), the findings of which have been mixed. While an episodic memory factor has generally been found, earlier studies were primarily concerned with semantic memory and STM and did not include a variety of measures of episodic LTM.

More recently, Unsworth and colleagues have done several studies evaluating the relation of retrieval from LTM and its relation to different memory systems as well as its role as a predictor of general intelligence (Unsworth, 2009; 2010; 2016; Unsworth, Fukuda, Awh, & Vogel, 2014). Unlike much other research evaluating the structure of memory abilities, Unsworth (2010) is notable in that it assessed episodic LTM in diverse ways. These included four different recall tasks, including free recall, cued recall, and source recall tasks in which participants had to remember items from a specific list. In addition, there were three different recognition tasks, all of which tested both item and source recognition, as well as five measures of working memory. The Unsworth's memory measures primarily used word lists, although one of the recognition measures assessed memory for picture items and their source (i.e., their location) and one of the working memory measures used visuospatial items. Analysis using CFA found that although they shared considerable variance, recall, recognition, and working memory were all separate abilities. However, the limited number of nonverbal measures precluded assessing the possibility that there are distinct nonverbal memory factors, and the possibility of a separate source memory factor was not considered.

1.1.4 Current Study

The goal of the current study was to explore the latent structure of episodic LTM measures. Participants were asked to recall or recognize items from different list-types, as well as the contexts in which the items were presented in. This allowed for an evaluation of source memory in its relation to recall and recognition. Source memory is generally measured in a two-step fashion, where participants are first asked if they recognize an item, and then asked to recall or recognize the context it was presented in. In this study they were directly asked to either recall or recognize the context of items, allowing for evaluation of the role of source memory directly with the same number of items per person. Note that the term context memory is used to describe the tasks used in this study, since the context is what the participants were asked to remember, but the context tasks were designed to be tasks that fall within the source memory construct.

This study also differs from previous studies by using a variety of items beyond simply words. Items used included words, names, nonwords, and images of nameable objects, and were chosen to vary on the dimensions of familiarity, semantics, and visuality. Standard word lists are familiar items that are verbal and also have meaning, presumably tapping semantic LTM memory. Lists of common English names are also familiar verbal items, but lack the semantic information associated with words. Nonwords are a third verbal stimuli that are both unfamiliar and lack semantic information. Finally, nameable objects were used as familiar, but pictorial stimuli that are likely to be encoded verbally (or at least can retrieved verbally). These different item types were used to make it possible to assess whether they tap different memory abilities. Online participants were used in addition to college students who may have developed unique strategies for memorizing information that differ from the rest of the population.

In sum, this study was designed to uniquely determine the latent structure of episodic LTM by assessing both recall and recognition, as well as memory for item and context, and by doing so using a variety of different types of items. Due to the relative novelty of some of the measures used, this study is primarily exploratory, but was designed to offer unique insights into the structure of memory abilities, which could eventually lead to a better understanding of the retrieval from LTM, how it is affected by aging, and its role in fluid intelligence. The tasks were designed to answer two questions; whether there are separate abilities for recall and recognition memory and whether there are separate abilities for item and context memory.

1.2 Method

1.2.1 Participants

Participants were recruited from two sources, Introductory Psychology students at Washington University in St. Louis and young adults from Amazon's Mechanical Turk (MTurk). Introductory Psychology students participated for partial course credit, while MTurk participants were paid $4.00 for their participation, which took between 30-40 minutes on average. MTurk workers have been demonstrated to be more diverse than traditional samples (Burmeister, Kwang, & Gosling, 201; Chandler & Shapiro, 2016) and still provide similar quality data as laboratory samples (Paolacci, Chandler, & Ipeirotis, 2010). In total 134 participants from MTurk and 113 students from Washington University in St. Louis completed the tasks. Due to the prevalence of problematic responding in online surveys, exclusion criteria were set up prior to any analyses. All participants who responded in less than 100ms on greater than 25 of the test trials (trials in which a memory decision had to be made) or responded inconsistently with the prompt on free recall trials (e.g., writing irrelevant sentences instead of typing single words) were excluded. Applying these criteria caused data from 84 participants (38 MTurk) to be removed, leaving a total of 164 participants (90 females) aged between 18 and 26 (M = 21.66, SD = 3.06) for analysis.

Participants recruited via MTurk were significantly older, $t(72.37) = 4.10$, $p < .001$, and more educated, $t(136.77) = 9.57$, $p < .001$, than the undergraduate participants, unsurprisingly given that the Introductory Psychology student participants were primarily first year students. Complete demographic information is presented in Table 1.

Table 1
Demographic Characteristics

	Age	Education (years)	Gender			Race		Ethnicity	
					Introductory Psychology Students				
Mean (SD)	18.51 (.68)	12.67 (.80)	Count (%)	Female	42 (60.9)	White/Caucasian	48 (69.6)	Hispanic/Latino	9(13.0)
Min	18	12		Male	27 (40.1)	Black/African American	5 (7.2)	Non-Hispanic/Latino	60(87.0)
Max	21	15				Asian	4 (5.8)		
						Multiple	9 (13)		
						Prefer Not to Respond	3 (4.3)		
					MTurk workers				
Mean (SD)	23.93 (1.85)	14.71 (1.85)	Count (%)	Female	44	White/Caucasian	67 (69.8)	Hispanic/Latino	10 (10.4)
Min	19	12		Male	44	Black/African American	11 11.5)	Non-Hispanic/Latino	86(89.6)
Max	26	20				Asian	6 (6.3)		
						Native American	1(1.0)		
						Multiple	9 (9.4)		
						Prefer Not to Respond	3 (4.3)		
					Total				
Mean (SD)	21.66 (3.06)	13.86 (1.81)	Count (%)	Female	90	White/Caucasian	115 (69.7)	Hispanic/Latino	19 (11.5%)
Min	18	12		Male	75	Black/African American	16 (9.7)	Non-Hispanic/Latino	146 (88.5%)
Max	26	20				Asian	10 (6.1)		
						Native American	1 (.6)		
						Multiple	18 (10.9)		
						Prefer Not to Respond	5 (3)		

1.2.2 Materials

All tasks were administered online on a webpage hosted on a private server. The tasks were programmed in HTML and JavaScript using JavaScript plugins from the library jsPsych (deLuuw, 2015).

Stimuli. There were four different types of stimuli (see samples in Figure 1).

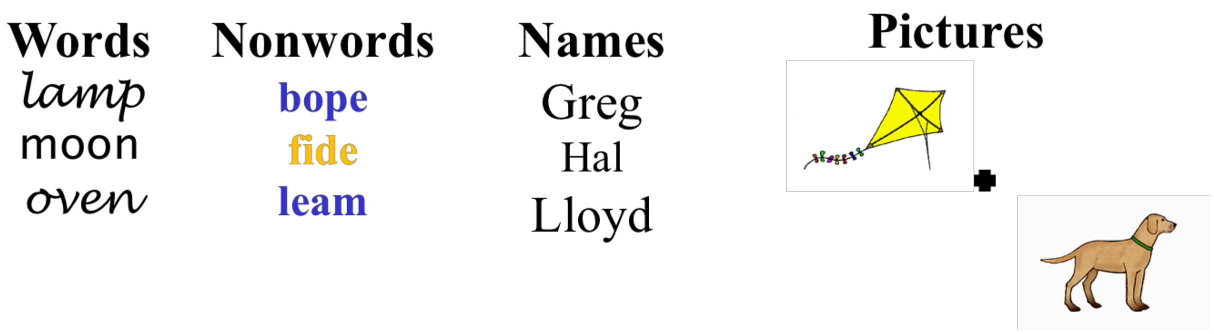


Figure 1. Sample Stimuli Presented in Context.

Words. Words were all single-syllable concrete nouns consisting of four or five letters. The words were selected from the English Lexicon project database (Balota et al., 2007) to be greater than average in terms of frequency in appearance, and behavioral performance also indicated better performance than average (i.e., fast RTs in lexical decision tasks, and high proportion correct on a naming task). Words were always presented for 1 second in the center of the screen in either script (Lucida Handwriting) or text (Courier New, a Serif font) font in all lowercase letters during the study phase, regardless of ultimate test-type. Neutral presentations (i.e., during test phases that were not context recognition) were in all-caps and in a sans-Serif font, (Lucida Console). Words were presented as image files in order to avoid differences in browsers handling of text font.

Nonwords. Nonwords were also single-syllable items consisting of four or five letters. The nonwords were created by changing one letter (usually the first letter) of a real English word (e.g., mice - bice). Nonwords were always presented for 1 second in the center of the screen in either blue or orange colored font during the study phase. Neutral presentations were in black.

Names. Names were one or two syllable common English names. Names were selected from published lists of the most common American baby names in various years, as published annually by the Social Security Administration. Names were consistent in each study list as taken from either common names for boys or girls. In order to avoid having names containing additional semantics, any names that were also common English words (e.g., Mark, Rose) were not used. Due to this limit, some two syllable names were used in order to reach 125 names that were still considered common. Names were presented in the study phase for 1 second in the center of the screen in either large (110 pixels) or small (30 pixels) font size, with the first letter always capitalized. Neutral presentations were in a medium font size (60 pixels), with the first letter capitalized.

Pictures. Images were taken from colored version of the classic Snodgrass pictures (Rossion & Pourtois, 2001). These pictures were selected since they were designed to be easily named. Pictures were presented for 3 seconds in one of four quadrants of the screen, with a fixation cross in the center of the screen to help clarify the quadrant the picture was in. During neutral presentations the pictures appeared in the middle of the screen.

Memory measures. Each set of stimuli were presented in four separate study lists of 25 items, with a different test following each list, for a total of 16 measures.

Item recall. Participants were instructed to type every item they remembered from the preceding list separated by a comma. Correct items were scored as exact correct spellings or spellings that were phonetically equivalent, and scores were proportion of the 25 items correctly recalled.

Item recognition. Each item from the original study list was presented intermixed with 25 new items. Items were shown in the neutral formats described above. After each item respondents were asked for their confidence on a scale of 1(low) to 3 (high) that the item was in the study list. Performance and confidence ratings were converted to receiver operating curves (ROCs) and scores were areas under the curve (AUC) for these ROCs. Based on pilot data where participants performed at ceiling on picture item recognition, the picture item recognition task was slightly different so that rather than having completely new items as foils, new items were versions of the same pictures that were altered by either adding or removing a part of the image (e.g., adding a hat onto an image of a clown). The list length for this task was 30 items rather than 25, and at test 15 items were the original items and 15 items were altered.

Context recall. Each item from the original study list was presented in neutral format. Participants were asked to choose which context the item was presented in (e.g., blue or orange). Scores were proportion correct.

Context recognition. Each item from the original study list was presented in either its original context or the other context (or one of the other three contexts in the case of the picture recognition). Confidence ratings were given and scores were calculated in the same way as item recognition tasks.

1.2.3 Procedure

Other than the crediting and payment process, the procedure was identical for both MTurk workers and undergraduate student participants. Once the website loaded, participants were given access to the information sheet containing the basic details of the study. After agreeing to participate, they answered brief demographic questions before the main test started. Prior to any of the memory tests, general instructions about the tests were given that outlined the different types of materials that participants would see and the tests they would be given after study. After the general instructions, the participants were given an indicator that the first study phase would begin and generic information about the type of material they would see. These instructions were the same regardless of which of the four tests the participants would take after the study list and participants were instructed to try to both remember the item and the context it was presented in. After the study phase, specific instructions on the type of test were given. To minimize the effects of different reading times a six second delay was added after participants pressed the spacebar key indicating they had read the instructions, meaning that the delay between study and test was 6 seconds plus the amount of time taken to read the test instructions.

All memory tasks were presented to the participants consecutively, in the same order for each participant. This task order was randomly selected, with the exception of the picture tasks, which were presented last, because pilot data suggested that the picture tasks may interfere with participants encoding and retrieval strategies for the verbal tasks. The task order was as follows: Word Item Recognition, name Item Recall , Word Context Recognition, Name Context Recall, nonword Context Recall, Word Item Recall, Word Context Recall, Nonword Item Recall, Name Item Recognition, Nonword Item Recognition, Name Context Recognition, Nonword Context Recognition, Picture Context Recognition, Picture Item Recall, Picture Context Recall, Picture Item Recognition.

After completion of all memory tasks participants were given a link to a debriefing form and given contact information for any questions and concerns, as well as a unique identifying code for credit or payment.

1.3 Results

The mTurk and undergraduate samples did not differ in mean score for any of the memory measures (all *t*s > 1.90, *p*s < .05), and the data from both samples were combined for all analyses. Complete descriptive statistics for all the measures can be seen in Table 2.

Table 2

Descriptive Statistics and Powers found for Box-cox Transformations for all Memory Measures

Measure	Mean	SD	Median	Range	Skew	Kurtosis	Power Transform
NW IT Recog	.67	.15	.70	.15-.91	-1.03	1.21	2.25
W_IT_Recog	.68	.18	.71	.21-.99	-0.52	-0.44	1.70
NM IT Recog	.72	.18	.76	.15-.99	-0.87	0.37	2.48
PIC IT Recog	.56	.12	.56	.25-.92	0.06	0.29	0.97
NW C Recog	.59	.13	.59	.27-.97	0.15	-0.28	0.62
WD C Recog	.62	.15	.62	.26-1.00	0.24	-0.26	0.87
NM C Recog	.70	.14	.70	.32-.99	-0.12	-0.68	0.83
PIC C Recog	.77	.16	.78	.37-1.00	-0.40	-0.83	1.25
NW C Recall	.54	.11	.52	.24-.88	0.25	-0.11	*NA*
WD C Recall	.54	.09	.52	.24-.76	-0.30	0.20	*NA*
NM C Recall	.65	.14	.64	.24-.96	-0.15	-0.04	1.95
PIC C Recall	.71	.22	.76	.12-1.00	-0.62	-0.65	0.97
NW IT Recall	.08	.07	.08	0.00-.48	2.18	7.81	*NA*
W IT Recall	.11	.11	.08	0.00-.64	1.88	5.16	0.38
NM IT Recall	.19	.11	.20	0.00-.68	1.01	2.18	0.60
PIC IT Recall	.24	.16	.20	000-.84	1.04	1.17	0.56

Note. NW IT Recog = Nonword item recognition; W IT Recog = Word item recognition; NM IT Recog = Name item recognition; Pic IT Recog = Picture item recognition; NW C Recog = Nonword context recognition; W C Recog = Word context recognition; NM C Recog = Name context recognition; Pic C Recog = Picture context recognition; NW C Recall = Nonword context recall; W C Recall = Word context recall; NM C Recall = Name context recall; Pic C Recall = Picture context recall; NW IT Recall = Nonword item recall; W IT Recall = Word item recall; NM IT Recall = Name item recall; Pic IT Recall = Picture item recall

1.3.1 Evaluation of Task Performance

Since source memory has traditionally been measured only with a single context recall or recognition measure, one question was whether the operationalizations used for recall for context and recognition for context resulted in tasks that truly differed, or if the tasks were essentially equivalent. For each type of stimuli, mean context recognition scores were greater than context recall scores(all *t*s > 4.00, *p*s < .001) suggesting that there was at least a difference in difficulty between the tasks and therefore that the question of whether the tasks rely on different abilities (the primary interest of this study) was worth further exploration.

As can be seen in Table 2, mean performance was barely above chance in the Nonword Context Recall (M = .54, Mdn = .52) and the Word Context Recall (M = .54, Mdn = .52) tasks. Given that most participants were performing at approximately chance, attempting to explain the variance in these tasks would be largely capitalizing on noise, so data from this task were removed from all further analyses. The Nonword Item Recall task also was removed because of limited variability as participants struggled with this task and were only able to recall two nonwords on average. In addition, participants frequently wrote answers that contained some part of the target words but were not phonetically equivalent. Notably, there was no *a priori* rule for scoring these items nor was there any consistent way to score them in a manner that accurately assessed the memory of the participant for the actual item.[1] After exclusion of these three tasks, there were seven participants who scored at chance or lower on more than six tasks. This constituted over half of the remaining 13 tasks, therefore, these seven participants were excluded from the present analyses.[2]

Only one remaining task, Word Item Recall, had univariate values of skewness or kurtosis above the generally accepted values (> 2; West et al., 1996). However, the combined dataset did violate multivariate normality (Mardia's Skewness = 1061.52, $p < .001$, Mardia's Kurtosis = 7.02, $p < .001$). Multivariate normality is not assumed for most exploratory factor analysis techniques. When variables violate multivariate normality, however, it can lead to

[1] When included in the analysis, nonword item recall had a very low communality (h^2= .04), indicating that almost none of the variance of the measure was explained. Including it in the analysis did not meaningfully change the overall factor structure.

[2] Analyses were run with all participants included and overall the same pattern of results were found. Additionally, a strict exclusion criteria where all participants who scored at chance on greater than 1 task was run and again the pattern of results was largely the same, despite the small sample size (n =120).

incorrect findings (Cain, Yang, &, Zhang, 2013). Furthermore, visual inspection of distribution plots revealed measures with differently shaped distributions, which could minimize the associations found between measures. In order to avoid these potential problems, all scores were power-transformed using Box-Cox techniques (Box & Cox, 1964) to best meet multivariate normality (see Table 2 for the powers used). After transformation, multivariate normality was no longer violated (Mardia's Skewness = 502.40, p = .06; Mardia's Kurtosis = 0.29, p = .77). Correlations of the transformed scores, shown in Table 3, were all positive and generally moderate in size.

1.3.2 Exploratory Factor Analysis

An exploratory factor analysis (EFA) was conducted using the R package "psych" (Revelle, 2013) in order to understand the latent variable structure of the memory measures. There are several different approaches available to determine the number of factors to extract for EFA. Courtney (2013) recommended five procedures, two of which were used here: Parallel Analysis and Velicers Minimum Average Partial (MAP) test. Parallel Analysis generates random matrices with the same number of cases and observations as the observed data and keeping only the number of factors whose eigenvalues are greater than those found in the randomly generated matrices is recommended. Figure 2 shows the eigenvalues obtained versus those found in the randomly generated data, and the plot suggests extracting four factors. Meanwhile, the MAP test calculates the average squared off-diagonal correlations from a matrix with an increase in the number of factors that were partialled out. The minimum average partial correlation is considered the best fit. The MAP procedure for these data (see Figure 3) suggests extracting two factors. Although the MAP procedure may be slightly biased towards underextraction (Ruscio and Roche, 2012), it is often best practice to be conservative in exploratory analyses. Therefore, the two-factor solution was chosen as the most likely to have a stable solution

Table 3

Correlations between all Memory Measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1. NW IT Recog	-												
2. W_IT_Recog	.57	-											
3. NM IT Recog	.49	.67	-										
4. PIC IT Recog	.22	.28	.31	-									
5. NW C Recog	.03	.06	.17	.25	-								
6. WD C Recog	.24	.32	.35	.23	.39	-							
7. NM C Recog	.26	.20	.36	.13	.02	.14	-						
8. PIC C Recog	.33	.35	.42	.37	.18	.32	.25	-					
9. NM C Recall	.26	.31	.28	.23	.26	.42	.15	.21	-				
10. PIC C Recall	.24	.31	.43	.43	.13	.24	.25	.67	.25	-			
11. W IT Recall	.24	.34	.41	.26	.19	.35	.13	.32	.30	.41	-		
12. NM IT Recall	.18	.20	.34	.18	.05	.17	.12	.36	.18	.41	.47	-	
13.PIC IT Recall	.27	.28	.35	.29	.20	.34	.09	.45	.25	.50	.53	.51	-

Note. NW IT Recog = Nonword item recognition; W IT Recog = Word item recognition; NM IT Recog = Name item recognition; Pic IT Recog = Picture item recognition; NW C Recog = Nonword context recognition; W C Recog = Word context recognition; NM C Recog = Name context recognition; Pic C Recog = Picture context recognition; NW C Recall = Nonword context recall; W C Recall = Word context recall; NM C Recall = Name context recall; Pic C Reall = Picture context recall; NW IT Recall = Nonword item recall; W IT Recall = Word item recall; NM IT Recall = Name item recall; Pic IT Recall = Picture item recall. All correlations greater than .15 are significant ($p < .05$).

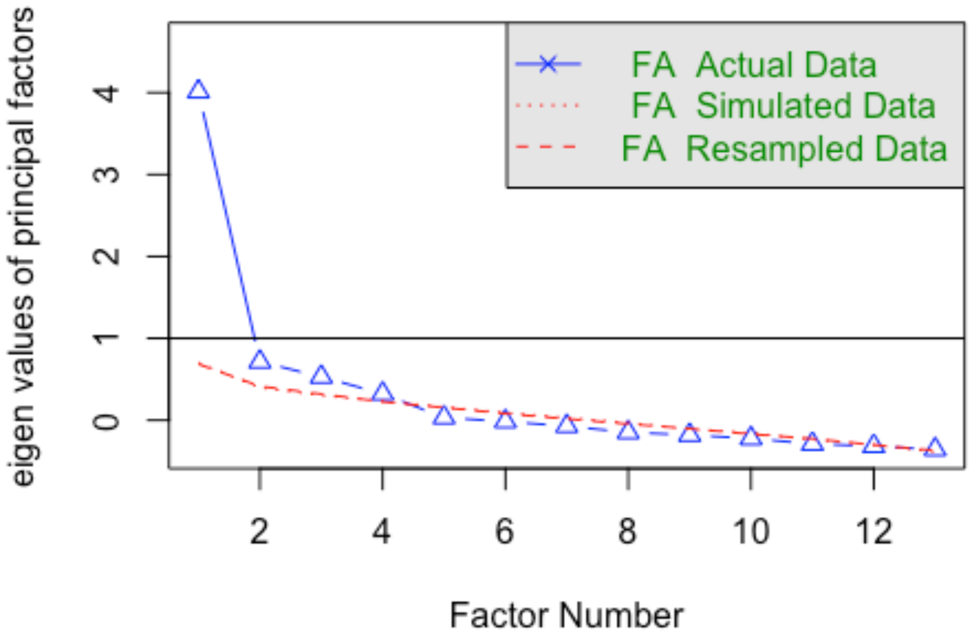
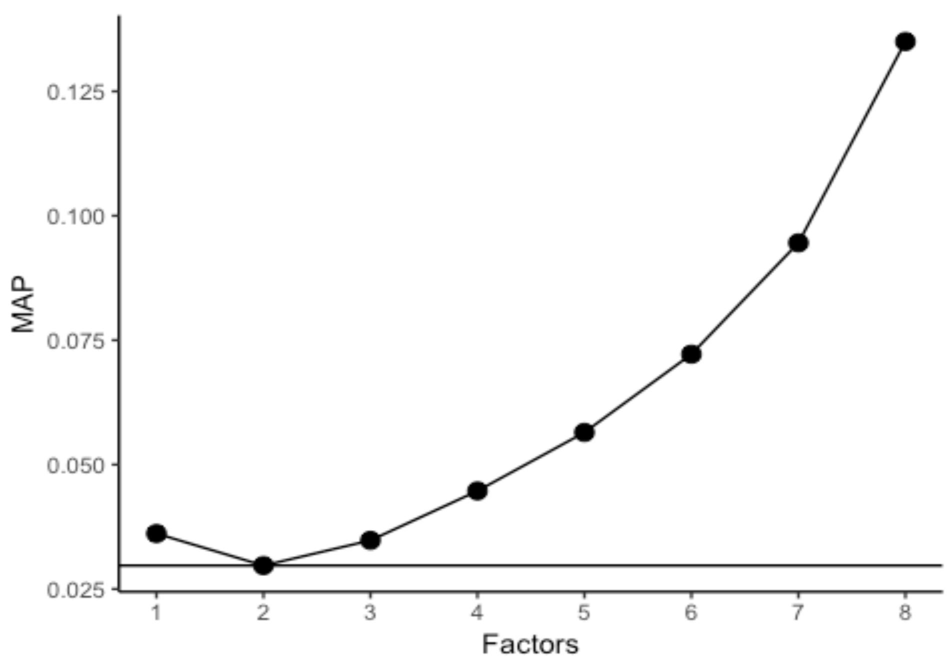


Figure 2. Scree Plots with Parallel Analysis. Parallel Analysis suggested a four-factor solution would best explain the data, as the actual eigenvalues exceeded the eigenvalues from resampled data for the first four factors extracted.



.

Figure 3. MAP Index as a Function of Factors. Minimum value is found with a two-factor solution.

The factors were extracted using maximum likelihood estimation. Although maximum likelihood relies on the assumption of multivariate normality, it should be noted that that assumption was no longer violated with the transformed data. Using an alternative procedure (ordinary least squares to find the minimum residual solution) produced an essentially equivalent solution. Factors were obliquely rotated using oblimin rotation. An oblique solution was preferred over orthogonal rotation because the factor correlations were extremely high (r = .54). As can be seen from the structure matrix presented in Table 4, the first factor was a general factor on which most of the tasks had relatively high loadings, particularly the picture tasks. The second component was characterized by high loadings from the item recognition tasks, suggesting the presence of a separate but correlated recall and recognition latent factor for item memory. The context measures tended to have moderate to low loading on both factors, with the exception of picture context (location), which suggests that the context memory measures may not measure the same abilities as the item memory measures.

Although a conservative approach to exploratory factor analysis is often best, choice of the “correct” solution is subjective. The results of the four-factor solution are reported in Table 5, but it should be noted that while the third and fourth factors may seem informative, they ought to be interpreted with caution. Again, oblimin rotation was used, and all correlations between factors exceeded .30 (see Figure 4). Comparisons of the model fit statistics of the two and four factor solutions are presented in Table 6. While these statistics tend to favor the four-factor model, the issue is not which model fits these data best per se, but which model accurately reflects the true latent factor structure.

Similar to the two-factor solution, the four-factor solution had a factor for item recognition, which was now the factor that explained the most variance. Rather than a general factor, there were three additional factors. The factor which explained the second most variance was primarily a pictorial factor, particularly picture context (location) memory. The third factor consisted of primarily loadings from the item recall tasks, further suggesting a separation between item recognition and recall. The factor explaining the least variance consisted of primarily loading from the remaining context tasks (i.e., not picture), although notably not name context recognition.

Table 4

Factor Loadings From Structure Matrix and Communalities for Two Factor Solution

Measure	ML2	ML1	h^2
PIC_C_RECALL	**.78**	-.02	.59
PIC_IT_RECA	**.73**	-.04	.50
PIC_C_RECOG	**.66**	.08	.51
NM_IT_RECALL	**.62**	-.06	.35
W_IT_RECALL	**.55**	.11	.38
PIC_IT_RECOG	**.41**	.12	.24
W_C_RECOG	.32	.23	.23
NW_C_RECOG	.29	-.02	.08
W IT Recog	-.07	**.90**	.75
NM_IT_RECOG	.20	**.67**	.63
NW_IT_RECOG	.01	**.65**	.43
NM C Recog	.14	.25	.11
NM C Recall	.24	.24	.18

Note. Bold indicates loadings .40 or greater. Factor names are numbered (ML1, ML2) according to variance explained in unrotated solution (i.e., ML1 explains more variance than ML2). Position left-right on table is in order of variance explained by final, rotated solution (i.e., ML2 explains more variance than ML1).

Table 5

Factor Loadings From Structure Matrix for Four Factor Solution

Measure	ML2	ML1	ML4	ML3	h^2
W IT Recog	**.88**	-.02	-.02	-.01	.74
NM IT Recog	**.67**	.13	.09	.04	.63
NW IT Recog	**.66**	.00	.01	-.01	.44
NAM C Recog	.26	.25	-.12	.00	.14
PIC C Recall	-.03	**.91**	.04	-.04	.84
PIC C Recog	.10	**.67**	.00	.08	.57
PIC IT Recog	.10	**.40**	-.06	.20	.27
NM IT Recall	-.01	.03	**.74**	-.13	.52
PIC IT Recall	-.03	.12	**.64**	.11	.57
W_IT_Recall	.10	-.06	**.63**	.14	.52
W_C_Recog	.12	-.02	.08	**.65**	.54
NW_C_Recog	-.15	.05	-.03	**.64**	.37
NM_C_Recall	.16	.02	.07	**.42**	.26

Note. Bold indicates loadings .40 or greater. Factor names are numbered (ML1, ML2, etc.) according to variance explained in unrotated solution. Position left-right on table is in order of variance explained by final rotated solution.

Table 6
Model Fit Statistics for Two Factor and Four Factor Solutions

Model	Chi-squared	df	R^2	RMSEA	BIC
Two Factor	**115.28**	53	.41	.090	-153.04
Four Factor	**28.66**	32	.50	.000	-133.34

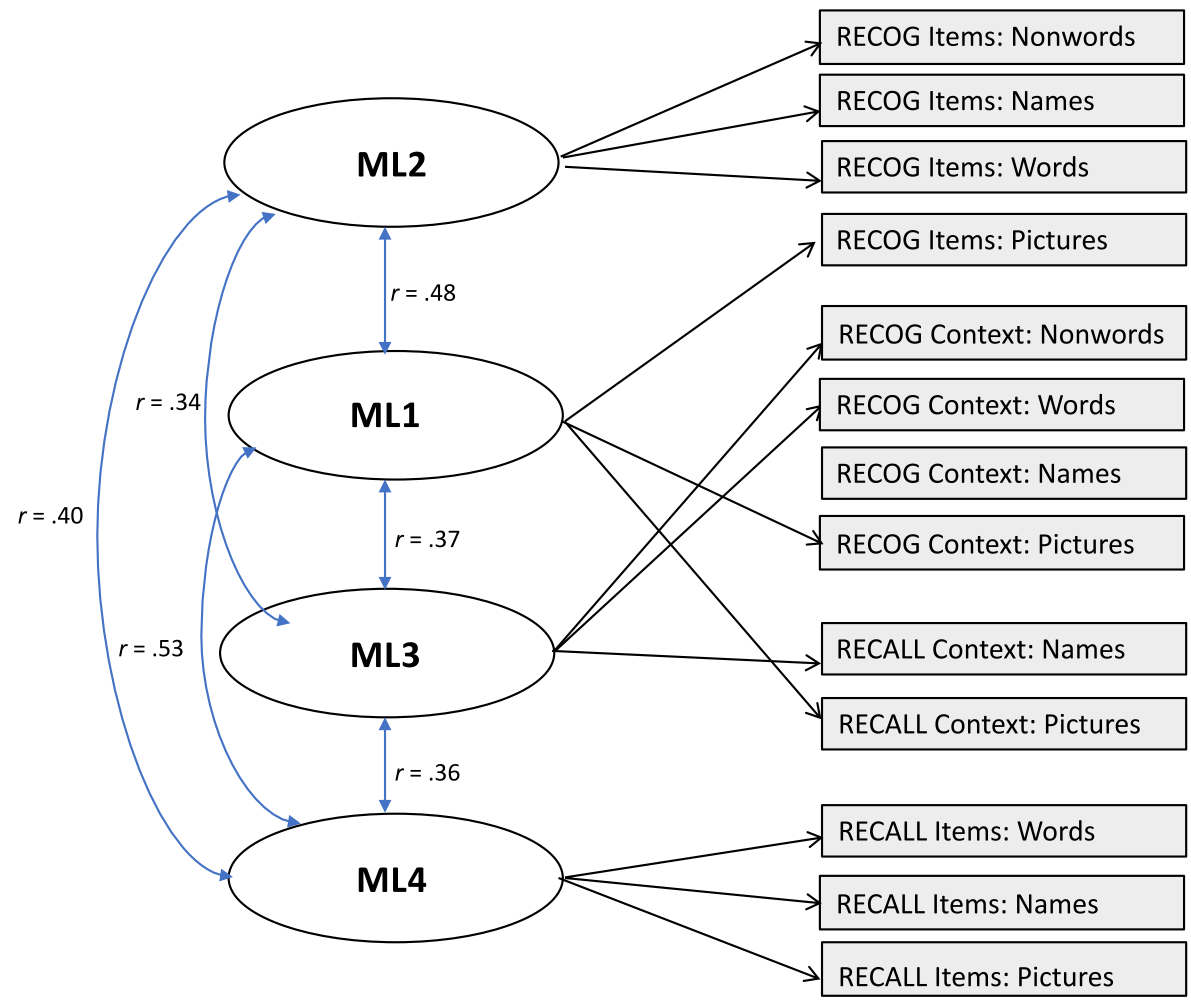


Figure 4. Path diagram of Four Factor Solution.

1.4 Discussion

The primary goal of this research was to evaluate the latent variable structure of various episodic LTM tasks. The tasks were designed to examine whether there are separate latent abilities for recall and recognition and for items and context (or source), using different types of material. In order to address the first question, whether recognition and recall are different abilities, there were five recall and eight recognition tasks that were evaluated in the analysis. Examination of a two-factor model using exploratory factor analysis found evidence that the three recognition tasks involving items presented orthographically (i.e., recognition of words, names, and nonwords) loaded on one factor, and word and name item recall loaded on another factor along with the four pictorial tasks, suggesting that memory for orthographic items relies on different abilities when recognizing than when recalling and that different abilities may be involved depending on whether items are pictorial (i.e., pictures or locations) or orthographic (i.e., names, words, or nonwords). The two factors in the solution were highly correlated, indicating that there is likely some general episodic LTM ability.

The finding that the verbal item recognition tasks loaded on their own factor suggests that it is not simply differences in measurement processes that separate these factors, as other tasks, namely context recognition and recall, relied on the same two alternative forced choice decision and did not load highly on the same factor. This suggests that an individual's ability to recognize verbal items from memory is a unique ability. Additionally, while item recall was not a separate factor in the two-factor model, when four factors were extracted there were separate recall factors for orthographic and pictorial items, on the one hand, as well as separate orthographic and pictorial recognition factors, on the other hand, further suggesting that item recognition and item recall involve separate abilities.

Although the answer to the question of whether recognition and recall involve separate abilities seems clear, the answer to the question of whether item memory is separable from context memory is much less so. While a separate factor on which only orthographic item recognition tasks loaded highly suggests that item memory and memory for context were at least partially different, there was little evidence for a separate context memory factor. In the two-factor solution, most of the context measures seemed to load modestly at best on both factors (*Mdn*s = .305 and .155 for the first and second factors, respectively) and only the pictorial context tasks loaded highly on a specific factor. Even allowing four factors did not reveal a factor composed of multiple context tasks all of which had high loadings; there was just one factor (ML3), which was the factor explaining the least variance, on which only context tasks loaded and one of those loadings was only .42.

The lack of a factor that consisted of high loadings for the context memory tasks is consistent with Siedlecki's (2005) findings on source memory. Yet as Johnson (2005) has argued, this does not indicate that source memory tasks are not useful tools. Rather, it suggests that source memory tasks recruit most of the same processes as other episodic memory tasks. Indeed, it is important to recognize that all episodic memory tasks depend on memory of context (e.g., what items were on a particular list that was just studied).

Regardless of whether a separate source memory ability exists, the orthographic context memory tasks in this study did not load highly on either of the two primary factors found, which suggests that these tasks do not depend on exactly the same set of abilities as item recall and recognition. In particular, the orthographic context memory tasks did not load highly on either factor in the two-factor model or on a separate factor (e.g., ML3) in the four-factor model, even though, on average, accuracy on name context recognition tasks and name context recall tasks was fairly high, suggesting that participants were able to do the tasks successfully. It is possible that these tasks would have been more related to the other orthographic item context tasks if performance on those tasks had not been generally less accurate, so that two of them had to be eliminated entirely. It is unclear, of course, whether relatively inaccurate performance on the other orthographic context tasks contributed to not finding a separate context memory factor or whether it simply led to lower loadings on the factors that were found.

Another question that arose in the analysis of the current data was whether there are different memory abilities for different kinds of items. The two-factor solution did not find evidence for any of the item-types being separate. However, it is noteworthy that the tasks with the three highest loadings on the first factor all required memory for pictures. This revealed itself as a separate factor (ML1) when four factors were extracted. Given that other research has found evidence for visual and verbal abilities on similar memory tasks (Hermann et al., 2001; Siedlecki, 2007; Weschler; 1997), it is possible that this third factor is more stable than one may normally when smaller factors are extracted in EFA. It should be noted, however, that the tasks used in this study were not well suited for examining a visual-verbal distinction because the only visual items were nameable objects, which may be represented verbally. Thus, although the current results should not be taken as strong evidence for separate visual and verbal episodic LTM abilities, they certainly do not rule out their existence and instead suggest that further research on this distinction might be informative.

The current study was primarily concerned with individual differences in LTM abilities, and not necessarily with distinctions between specific memory systems. However, the results can still be interpreted in the broader framework of theories of memory systems. The dual process theory of recognition, in which recollection and familiarity are the two processes that make up recognition and recall tasks, might explain the results of the study fairly well. That is, the present pattern of results is fairly consistent with individual differences in the efficiency of these processes that manifest themselves as one factor reflecting the ability to use recollection and another factor reflecting the ability to use familiarity. The correlation between the two factors that reflect these abilities makes sense because, according to the dual process theory, recognition relies on both recollection and familiarity, and thus the item recognition tasks would not be totally separable from the context memory tasks. That is, the context memory tasks may rely on recollection because in order to make context judgment, one must remember both the items and the specific contexts in which each item occurred, and thus may involve recollective experiences. If so, the fact that both kinds of tasks involve recollection would explain why the picture context tasks loaded on the same factor as item recall tasks.

This view is consistent with that of Yonelinas (1999), who argued that source memory tasks primarily rely on recollection although familiarity may also play a role depending on the exact nature of the task. The context recognition tasks could have relied more on familiarity than most context memory tasks, as participants could simply have had a feeling that an item looks familiar. Even the context recall tasks could rely somewhat on familiarity because participants did not necessarily have to entirely recollect the context in which an item appeared, but could choose the context that seemed more familiar (without actually seeing the item again in either context). In other words, the context tasks may differ slightly in how much they rely on recollection versus familiarity, leading to a less clean break between the two factors and lower loadings. These explanations are somewhat speculative as the current tasks were not designed to separate recollection and familiarity, and the fact that the recollection-familiarity framework may be able to explain these data does mean that other theories could not offer alternative explanations.

1.4.1 Limitations and Future Directions

The primary limitation of this study was that three out of the sixteen tasks had to be eliminated from the analysis due to low accuracy and some of the remaining tasks also suffered from relatively poor performance in that a substantial number of participants performed at chance or lower on those tasks. Three potential problems were identified that could explain chance performance; two having to do with the design of the tasks and one having to do with the overall study design. The first problem was that nonwords were more difficult than either words or names, and two of the tasks that were eliminated involved nonwords. As discussed earlier, scoring recall of nonwords was difficult because of differences with respect to the spelling of reported items. Additionally, participants seemed to have trouble recalling the contexts (i.e., the colors) in which nonwords appeared, and even context recognition was difficult (M = .59).

Future studies should consider using alternative items such as low frequency words in place of nonwords because unfamiliar words may be easier to remember and scoring recalled items would be more straightforward.

The second problem identified was that some contexts were less distinctive than others. The word context tasks were approximately as difficult as the nonword context tasks, most likely because the contexts used (i.e., the fonts) were probably not equally “memorable,” with words in script being more distinctive than words in a normal font. Theoretically, one could get a perfect score on the context tasks if one only remembered the words in script and assumed the others were in the normal font, but the participants had to try to remember all of the words as well as their context because they did not know whether they would be tested on the items or their context. Future studies should ensure that all the forms of context that are used are equally distinct, and in particular that the variations within contexts are equally distinct (e.g., size differences are as discriminable as color differences).

The third problem was the design of the online study. In online studies, researchers have less control over what the participants are doing and whether they are maintaining attention on the task. While lapses of attention undoubtedly occur in the laboratory as well, they can be monitored much more effectively than attention in online studies. This study presented the memory tasks in consecutive order with no built-in break or any tasks that differed in a significant way. As a result, participants may have gotten bored performing the memory tasks due to their repetitive nature. Note that this was not specific to MTurk workers, as MTurk workers did not score significantly lower than the undergraduate psychology students on any task, and the MTurk workers and undergraduate samples did not differ in the percentage excluded due to problematic responding. Future memory studies should build in breaks, especially for online participants, and try to mix up the types of tasks that participants are asked to complete, as well as add possible incentives for performing tasks well.

Although the current study was concerned with the structure of memory abilities such as might be revealed with a latent variable approach, EFA was used because of issues with chance performance and the novelty of some of the tasks, including the ways in which context memory was measured. CFA is the latent variable approach preferred by most researchers because it allows for direct comparison of models that are consistent with different theories. Accordingly, future research should use CFA to better understand the nature of episodic LTM abilities. Using CFA and structural equation modeling can also be beneficial in understanding the relation of LTM abilities to higher order cognitive abilities such as fluid intelligence. Studies such as Unsworth (2010) have demonstrated that LTM does not predict fluid intelligence beyond its association with working memory, but examination of more complex models of LTM abilities may lead to different conclusions.

1.4.2 Conclusion

The current study evaluated the distinction between episodic LTM tasks from an individual differences perspective. A range of different LTM tasks that differed with respect to the kinds of to-be remembered items and whether the tasks tested participants' recall or their recognition of items or the context in which they were presented. The current results suggest that episodic LTM relies on at least two latent variables, one reflecting an ability specific to item recognition and the other reflecting a more general episodic LTM ability. This finding suggests that there is a difference between item recall and recognition but that there are not further distinctions among LTM abilities. Nevertheless, it should be noted that there was some evidence for a visual-verbal distinction, instantiated as distinction between orthographically presented items and pictorial (albeit nameable) items, although importantly, evidence for a separate context memory factor was not found. Future research that uses an approach conceptually similar to that used here but that improves on the specific tasks and analytic methods should help further our understanding of episodic LTM abilities.

References

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*(3), 563–580. https://doi.org/10.1016/S0749-596X(02)00502-8

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex*, *12*(10), 1048-1056.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Courtney, M. G. R. (2013). Determining the Number of Factors to Retain in EFA&58; Using the SPSS R-Menu v2. 0 to Make More Judicious Estimations. *Practical Assessment*, *18*(8), 1-14.

Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in Brain Research* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1-12. doi:10.3758/s13428-014-0458-y.

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion*, *1*(1), 5–11. https://doi.org/10.1037/1528-3542.1.1.5

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology*, *9*(2), 229.

Hale, S., Rose, N. S., Myerson, J., Strube, M. J., Sommers, M., Tye-Murray, N., & Spehar, B. (2011). The structure of working memory abilities across the adult life span. *Psychology and Aging*, *26*(1), 92–110. https://doi.org/10.1037/a0021483

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1195–1209. https://doi.org/10.1037/0278-7393.25.5.1195

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

James, William (1890), *The Principles of Psychology*, Vol. 1. New York: Holt.

Johnson, M. K. (2005). The relation between source memory and episodic memory: Comment on Siedlecki et al. (2005). *Psychology and Aging*, *20*(3), 529–531. https://doi.org/10.1037/0882-7974.20.3.529

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin*, *114*(1), 3.

Kausler, D. H., & Puckett, J. M. (1980). Adult age differences in recognition memory for a nonsemantic attribute. *Experimental Aging Research*, *6*(4), 349–355. https://doi.org/10.1080/03610738008258370

Kintsch, W. (1970). *Learning, memory, and conceptual processes.* Oxford, England: John Wiley

& Sons.

Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 31.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory?. *Psychological bulletin*, *135*(4), 638.

Nyberg, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 485.

Paivio, A. (1971). Imagery and language. In *Imagery* (pp. 7-32).

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psychVersion = 1.8.10.

Rossion, B., & Pourtois, G. (2001). Revisiting Snodgrass and Vanderwart's object database: Color and texture improve object recognition. *Journal of Vision*, *1*(3), 413-413.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282–292. https://doi.org/10.1037/a0025697

Siedlecki, K. L. (2007). Investigating the structure and age invariance of episodic memory across the adult lifespan. *Psychology and Aging*, *22*(2), 251–268. https://doi.org/10.1037/0882-7974.22.2.251

Siedlecki, K. L., Salthouse, T. A., & Berish, D. E. (2005). Is there anything special about the aging of source memory?. *Psychology and Aging*, *20*(1), 19.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*, 381–403.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Unsworth, N. (2009). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition*, *37*(6), 837–849. https://doi.org/10.3758/MC.37.6.837

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*(1), 16–28. https://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 50.

Wechsler, D. (1997). *Wechsler Memory Scale—III*. San Antonio, TX: The Psychological Corporation.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: a formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(6), 1415–1434.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London. Series B:*

Biological Sciences. Retrieved from https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2001.0939

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In *The handbook of aging and cognition, 2nd ed* (pp. 293–357). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.