Psychology and Aging
2007, Vol. 22, No. 2, 251–268

0882-7974/07/$12.00 DOI: 10.1037/0882-7974.22.2.251

Investigating the Structure and Age Invariance of Episodic Memory Across the Adult Lifespan

Karen L. Siedlecki
University of Virginia

The structure of episodic memory was investigated by assessing different modalities of material (verbal, figural, and spatial) and different types of tests (recall, cued recall, and recognition). A 3-factor model that distinguished among modalities of material was found to be the best representation of memory and the verbal, figural, and spatial memory factors exhibiting construct validity. This 3-factor modality of material model also demonstrated configural, metric, and structural age invariance across a sample of adults ($N = 327$) between the ages of 18 and 94. There was evidence that latent constructs corresponding to recall, cued recall, and recognition could be distinguished from one another within the verbal domain but not within the figural and spatial domains. A mediation model examining the retrieval constructs was examined within the verbal domain, and there were unique age-related influences on cued recall and recall performance. This result is consistent with findings that increased age is associated with increased difficulty in retrieving information.

Keywords: episodic memory, aging, structural equation modeling, invariance

Memory loss is the complaint most often associated with increased age (see Jonker, Geerlings, & Schmand, 2000, for a review), and there is considerable evidence indicating that episodic memory declines with age (see Zacks, Hasher, & Li, 1999, for a review). *Episodic memory* is defined as memory for information, facts, or events that have a time or place associated with them. Although memory has been the focus of thousands of research articles, the structure of memory, and its relation to age, is still unclear.

There are three main approaches to studying memory: the cognitive approach, the neurobiological approach, and the correlational approach. Across these three methods, there is converging evidence that episodic memory can be distinguished with respect to types of test and modality of material. In this article, material modality is categorized in terms of verbal, figural, and spatial stimuli. However, there are many different "modalities" in which information can exist. For instance, stimuli may also be characterized, among other things, as olfactory, auditory, or tactile. Verbal, figural, and spatial memory are three of the more studied types of memory, and thus these three types of material are examined in the current project.

Verbal material is typically defined as information that can be stored phonetically and may include words, facts, and stories. As noted by Klauer and Zhao (2004), the distinction between figural and spatial tasks often corresponds to a difference between "what" and "where" information such that *figural memory* refers to the "what" aspect of information like patterns, shapes, and colors, or the features of objects. In contrast, *spatial memory* refers to the "where" aspect of memory and incorporates both location and movement information. It should be noted that figural memory is commonly referred to as *visual memory* in the literature, but in addition to referring to type of material, visual memory may also refer to the modality of presentation. Thus to avoid any confusion, the term *figural memory* is used throughout this article to refer to the type of stimuli used in the current study.

Researchers have partitioned memory into verbal, figural, and spatial components using the typical cognitive univariate approach by, for example, demonstrating a double dissociation between figural and spatial tasks (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Klauer & Zhao, 2004) and selective interference between verbal and figural memory (e.g., Salthouse, 1974). Within the neurobiological domain there is evidence that patients with unilateral temporal lobe epilepsy (TLE) demonstrate impairment specific to material modality such that patients with left TLE show selective impairment of verbal memory and those with right TLE show selective impairment of figural memory (Delaney, Rosen, Mattson, & Novelly, 1980; Hermann, Wyler, Richey, & Rea, 1987; Loring, Lee, Martin, & Meador, 1988; Milner, 1975; cf. Wilde et al., 2003). Studies that have used different brain imaging techniques, such as positron emission tomography (e.g., Kohler, Moscovitch, Winocur, Houle, & McIntosh, 1998), functional magnetic resonance imaging (e.g., Hayes, Ryan, Schnyer, & Nadel,

Karen L. Siedlecki, Department of Psychology, University of Virgina.
Data collection for this article was supported by National Institute on Aging Grant R01 AG19627 to Timothy Salthouse.

A longer version of this article served as my dissertation while at the University of Virginia, and I acknowledge my committee members Chad Dodson, Jack McArdle, Deborah Roach, and Timothy Salthouse for their thoughtful suggestions on earlier versions of this article. In particular, I thank my advisor, Timothy Salthouse, whose advice, wisdom, and guidance have been invaluable. I also extend my gratitude to Cris Rabaglia and the Salthouse Cognitive Aging Lab research assistants for all their help during the data collection process.
Correspondence concerning this article should be addressed to Karen L. Siedlecki, who is now at Columbia University College of Physicians and Surgeons, Cognitive Neuroscience Division, Taub Institute, P & S Box 16, 630 West 168th Street, New York, NY 10032. E-mail: ks2513@columbia.edu

2004), and event-related potentials (e.g., Mecklinger & Meinshausen., 1998) have reported differences in brain regions activated during figural and spatial tasks.

Correlational studies provide additional evidence that verbal memory, figural memory, and spatial memory are distinct constructs (Elfkides et al., 2002; Herrmann et al., 2000; Wechsler, 1997a). For example, using the standardization sample of the Wechsler Memory Scale—III (WMS–III; Wechsler, 1997c), Wechsler (1997a) conducted confirmatory factor analyses on three age groups (16–29, 30–64, 65–89 years of age) and found that models that distinguished between verbal and visual memory fit best across all groups.

Type of test refers to the division among recall, cued recall, and recognition tests. There is considerable evidence within the cognitive psychology literature suggesting that recall and recognition involve some distinct processes. For instance, certain phenomena, such as proactive and retroactive interference, are evident in word recall but appear less often in tests of word recognition (M. C. Anderson & Neely, 1996). In addition, age differences are more often observed, and are more pronounced, in recall and cued recall tasks than in recognition tasks (Harwood & Naylor, 1969; Parker, Landau, Whipple, & Schwartz, 2004; Schonfield & Robertson, 1966; Warrington & Sanders, 1970). There is also evidence within the neurobiological literature of a recall–recognition distinction. For example, different areas of the brain have been shown to be active in recall and recognition tests (e.g., Cabeza et al., 1997), and studies of amnesic individuals have demonstrated that when compared with normal controls, amnesic individuals' recognition performance is less affected than is their recall performance (Hirst et al., 1986; Hirst, Johnson, Phelps, & Volpe, 1988; but see Haist, Shimamura, & Squire, 1992). Finally, correlational studies also provide evidence that performance on recall, cued recall, and recognition tests are separable in terms of patterns of individual differences (Nyberg et al., 2003; Underwood, Boruch, & Malmi, 1978).

The above (brief) review of the relevant research indicates that there are many facets to memory; in order to attempt an inclusive investigation of episodic memory, multiple aspects of memory need to be incorporated. Thus, in this project verbal memory was assessed by recall, cued recall, and recognition tests of lists of words. Memory for spatial information was assessed with recall, cued recall, and recognition tests of locations of circles within a grid. To measure figural memory, participants completed recall, cued recall, and recognition tests of line drawings. Figure 1 depicts the six potential latent constructs derived from the nine manifest variables measured in this project. This type of model is consistent with the multitrait–multimethod (MTMM) analytical method first articulated by Campbell and Fiske (1959) as a way to examine convergent and discriminant validity, and Widaman (1992) has argued for its use in studies of aging. However, it should be noted that in this study the models are tested independently without applying an MTMM analysis to the data.

There are three main goals of the study. The first is to determine the most meaningful model of episodic memory with a particular set of variables. The meaningfulness of the model is ascertained not only by the fit to the data but also by examining whether the model exhibits construct validity. Construct validity is determined by examining whether the variables hypothesized to represent a construct have a substantial amount of variance in common (i.e., convergent validity) and by examining whether the constructs hypothesized to represent distinct dimensions of individual differences do not have very high correlations with one another (i.e., discriminant validity). Thus for a construct to exhibit convergent validity, the standardized coefficients from the latent construct to the variable should be moderately high; and to exhibit discriminant validity, the correlations among the latent constructs should be relatively low (or at least significantly less than 1.0).

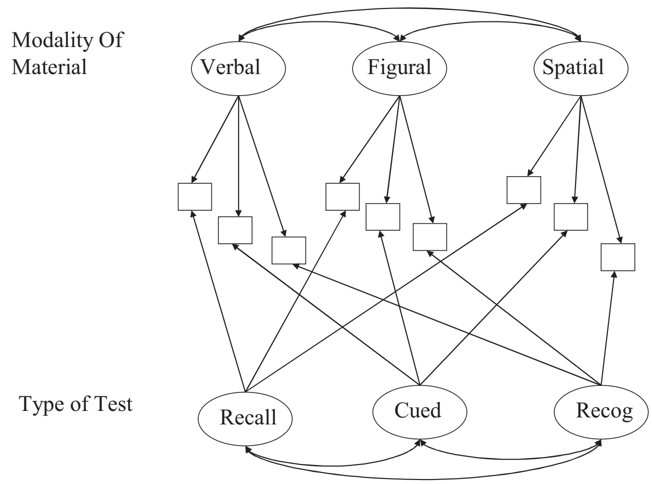

Figure 1. Simultaneous structural equation model of the nine manifest variables and the six hypothesized latent variables. Recog = recognition.

The second goal is to determine whether the most meaningful structure of memory (as determined by the goodness-of-fit indices for the set of theoretically driven models) is invariant across age. Horn and McArdle (1992) argued that if the structure of a construct differs across groups or conditions, then the "basis for drawing scientific inference is severely lacking; findings of differences between individuals and groups cannot be unambiguously interpreted" (p. 117). Two specific questions regarding invariance were addressed in this article. First, are the memory variables measuring the relevant constructs to the same degree across the three age groups? Second, are the relations among the factors invariant across the age groups? Of greatest interest to this project are the relations from the latent constructs to the observed variables across age to determine whether the variables are measuring the construct to the same degree. Researchers studying age-related differences in memory often compare the performance of a young and an older group of participants by conducting analyses of variances. An assumption intrinsic to this approach is that the differences between the groups are quantitative, and not qualitative. An important aspect of this project focuses on investigating this assumption with empirical data by examining whether the structure of episodic memory is the same across age groups. In one of the only studies to explicitly examine the invariance of memory across age, Nyberg et al. (2003) reported that a model of memory that distinguished between two higher order factors of episodic and semantic memory and among four lower order factors of recall, recognition, knowledge, and fluency exhibited invariance across three age groups ranging from 35 to 80. However, the youngest participants in the Nyberg et al. study were 35 years of age, which is considerably older than the college-age students used in most memory research. This study examines invariance of memory across a sample ranging from 18 to 94 years of age.

If a construct does not exhibit invariance, it suggests that the meaning of the construct is shifting across groups. In this case, quantitative comparisons across the groups are inadvisable. For example, Schaie (1996) reported on a vocabulary test designed for children, which when administered to adults was found to be a stronger measure of perceptual speed (presumably because the test items become too easy with increased age, and success is reflected by the quantity rather than the quality of answers in a given time period). In memory research, the comparison of older and younger groups assumes that the meaning of the memory construct is not shifting, and the differences are therefore a reflection of differences in performance (i.e., quantitative differences). Age invariance analyses are one way to ensure that the meaning of the construct of interest is not changing across groups.

The second question regarding invariance (i.e., are the relations among the latent constructs invariant?) is relevant because it has been hypothesized that as adults grow older, their cognitive abilities become dedifferentiated such that the constructs become more highly correlated with one another across age (e.g., Baltes, Cornelius, Nesselroade, & Willis, 1980). One way to examine the dedifferentiation hypothesis within the context of memory constructs is to examine the invariance of the correlations among the factors across age.

A third goal of the project is to examine the retrieval of information across the adult portion of the lifespan. Previous research has attempted to examine whether age-related memory difficulties occur in the encoding, storage, or retrieval stage (e.g., Park, Smith, Dudley, & Lafronza, 1989). *Encoding* refers to the process of acquiring information; *storage* refers to the process or stage of retaining information; and *retrieval* refers to the process of accessing, or retrieving, the information from memory.

Several different types of evidence suggest that increased age is associated with difficulty in the retrieval process. For instance, the tip-of-the-tongue (TOT) state is described as a state in which an individual has difficulty in retrieving a word, although the word can often be defined, and the first letter or even the number of syllables identified. In a diary study (Burke, MacKay, Worthley, & Wade, 1991), the TOT state was resolved and the appropriate word was retrieved in 95% of the cases, implying that the word was available and the difficulty was in retrieving the word. In two diary studies (Burke et al., 1991; Cohen & Faulkner, 1986) and in a study in which TOT states were experimentally induced (Burke et al., 1991), older adults have reported more TOTs than have younger adults.

In addition, there is evidence from studies of implicit memory that information may be present and expressed without being consciously retrieved. *Implicit memory* refers to the beneficial effects of past experience on a subsequent test. For example, in a word-stem completion task participants are more likely to fill in a word stem with a word they viewed previously even though the word-stem task was administered independent of the original task in which the words were presented and participants were not instructed to recall the words previously presented. Studies have shown that when information is not consciously retrieved there may be few to no age differences on measures of implicit memory (e.g., Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004; Light & Singh, 1987; Mitchell & Bruss, 2003; but see La Voie & Light, 1994, for a meta analysis examining cross-sectional data), suggesting that it is the act of retrieving that creates problems for older adults.

Although researchers investigating episodic memory differences across the adult lifespan have almost invariably reported that episodic memory performance declines with age, different tasks and different types of memory tests have been found to have different age relations. More specifically, measures of cued recall (Craik & McDowd, 1987) and recall (Nyberg et al., 2003; Parker et al., 2004) often have larger negative age relations than do measures of recognition memory. It is possible that some of the differences in the age relations may be due to differential reliability among recall and recognition tasks because a variable must be reliable in order to have systematic variance available to be related to other variables, such as age. Unfortunately, the reliabilities of these measures are seldom reported, and consequently it is unclear whether differential age relations exist because of true differences in the age association or because of differential reliabilities. However, it is still informative to examine substantive reasons behind the differential age relations to recall and recognition.

The differences among tests of recall and recognition have been a popular topic in memory research, and many theories exist to explain performance on these tests. An early view, a single-process theory of memory, maintained that there is one underlying memory system or process (e.g., memory strength), and because recall requires greater memory strength than does recognition, recall tests are more difficult than recognition tests (Postman, 1963). Although parsimonious, the single-process theories have trouble explaining many findings. A dual-process theory of recall emerged in the 1970s (J. R. Anderson & Bower, 1972; Kintsch, 1970), labeled the *generate–recognize model* of recall. In this model, recall is postulated to consist of two stages comprised of a generate stage and a recognize stage. In the generate stage, items are retrieved from memory that were potentially presented during the study phase. The recognize stage consists of a decision process, whereby studied items are distinguished from nonstudied items. This model proposed that recognition consisted of only the recognize–decision stage. The generate–recognize model has been revised to include such functions as control processes (e.g., Haist et al., 1992; Squire & Zola, 1998), monitoring (Moscovitch, 1994), and inhibition (e.g., Shimamura, 1995). Finally, dual-process theories of recognition postulate that the processes of familiarity and recall contribute to recognition judgments, whereas free recall relies primarily on recollection processes (e.g., Atkinson & Juola, 1973; Jacoby, 1991; see Yonelinas, 2002, for a review of research examining recollection and familiarity in recognition memory).[1]

[1] *Familiarity* has been conceptualized in a number of different ways: as an activation of lexical nodes representing words or objects (Atkinson, Hertmann, & Wescourt, 1974; Atkinson & Juola, 1973; cf. Yonelinas, 2002), as an assessment of processing fluency (e.g., Jacoby, 1991), or as an assessment of quantitative memory strength (Yonelinas, 2002). *Recollection* is typically conceptualized as the recovery of information about the to-be-remembered item (e.g., Jacoby, 1991) and as the process by which qualitative information regarding a studied item is retrieved (Yonelinas, 2002). Familiarity is often considered to be a faster, more automatic process, whereas recollection is considered to be a slower and more controlled process.

Consequently, one way to conceptualize the difference between recognition and recall is to separate the two test types in terms of hypothesized processes. For example, it was hypothesized that recognition includes the process of encoding and storage but only minimal retrieval processing. Because alternatives are presented to the individual in a recognition test, a retrieval process is not needed to generate the test item. Instead, there is evidence that though recognition relies on both familiarity and recollective processes (e.g., Yonelinas & Levy, 2002), recall relies primarily on more demanding recollective processes.[2] Because recall also includes encoding and storage processes, it can be hypothesized that the largest difference between recognition and recall is that recall requires retrieval to generate an answer.

Tests of cued recall are presumably intermediate between tests of recognition and recall because cued recall tests involve some retrieval processes. Given that there is an available cue, the retrieval effort should fall somewhere between a recognition test in which the response alternatives are provided and recall in which the complete response needs to be selected and generated. This argument is supported by Craik, Govoni, Naveh-Benjamin, and Anderson's (1996) findings that a secondary task during retrieval affected recall performance the most, affected recognition performance the least, and affected cued recall performance intermediately.

It has been argued that recall is particularly affected by age because increased age is associated with fewer processing resources (e.g., Craik, 1986; Craik & McDowd, 1987) to manage the increased retrieval demands inherent in recall tasks and to carry out self-initiated processing. Craik and colleagues (Craik, 1986; Craik & Jennings, 1992) have argued that environmental support (which is strong in cued recall and in recognition tests but weak in recall tasks) decreases the need for self-initiated processing.

The hypothesized relation between age and the types of memory tests is depicted in Figure 2. The relation from age to recognition is represented by a dotted line to signify that a relation exists but that it may be minimal. The relation from age to cued recall is represented by a solid line to indicate that the relationship was hypothesized to be greater in magnitude than the relationship between age and recognition. Finally, the relation from age to recall is represented by a bold solid line to indicate that the relation was hypothesized to be of the greatest magnitude.

A conceptually simple way to investigate this hypothesis is to examine Figure 2 in terms of a mediation model. If cued recall is more age-sensitive than is recognition, then there should be statistically significant age-related effects on cued recall that are independent of the age-related effects on recognition. That is, after partialling out the age-related effect on recognition, there should be unique age-related effects on cued recall. In addition, if recall requires additional retrieval processes above and beyond that of cued recall, then there should be unique and statistically significant age-related effects on recall after partialling out the age-related effect on cued recall.

In summary, there are three goals of this project. The first goal of the project is to use a correlational approach to determine a meaningful representation of the structure of episodic memory. Specifically, which theory-driven model provides the best fit to the data? To answer this question, two series of models are compared. The first set of models examine the constructs of verbal, figural, and spatial memory and whether a one-factor, a two-factor, or a three-factor model provides the best fit. The second set of models determine which model best describes the relations among the constructs of recall, cued recall, and recognition. The second goal of the project is to assess the age invariance of the model that is determined to be the best representation of memory (i.e., is the model invariant across age?). The third goal is to determine whether there are unique, statistically independent, age-related effects on cued recall after considering the influence of age on recognition and whether there are unique age-related effects on recall after partialling out the age-related effect on cued recall.

Method

Participants

Participants were 332 adults between 18 and 94 years of age recruited from the Charlottesville, VA, community via newspaper advertisements, flyers, and referrals from other participants. The participants reported to the laboratory three times for approximately 2 hr on each occasion, and each received $120 at the completion of the study. The tasks for the current study are a subset of those included in a larger project in which measures of core cognitive abilities such as vocabulary, processing speed, and fluid ability were also obtained.

The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) was administered, and 5 participants with scores below 24 were excluded from the subsequent analyses. Table 1 reports the characteristics of the resulting sample, divided into three age groups. Inspection of the table reveals that increased age was associated with lower self-ratings of health (although the average health assessment for the two older groups corresponds to a rating of "very good") and a greater degree of education. Scaled scores from the Wechsler Adult Intelligence Scale—III (WAIS–III; Wechsler, 1997b) and the WMS—III (Wechsler, 1997c) are reported for the Vocabulary, Digit Symbol, Logical Memory, and Word Recall subtests. The norms for this battery were created by using a nationally representative sample, and each subtest has a mean of 10 and a standard deviation of 3 in that normative sample. The average scaled scores for the current sample ranged from 10.9 to 13.0, indicating that the current sample was functioning at a level between one third and one standard deviation higher than the national average. Increased age was associated with higher scaled scores in both the Digit Symbol and Logical Memory subtests, suggesting that the older adults might be a more selective sample relative to their age group than might be younger participants, at least in regard to these two subtests.

Procedure

Figure 3 depicts how verbal memory, figural memory, and spatial memory were assessed with recognition, cued recall, and recall tasks.

[2] There are, of course, exceptions to this generalization. In some tests of recognition, recollective processes are primarily used. For example, Yonelinas (1997) administered recognition tests to participants and found both familiarity and recollection were used to recognize item information about a word (i.e., whether the word was old or new), as is typically reported. However, recollection was primarily used in recognition judgments regarding associative information (i.e., were these words paired together during the study phase?).

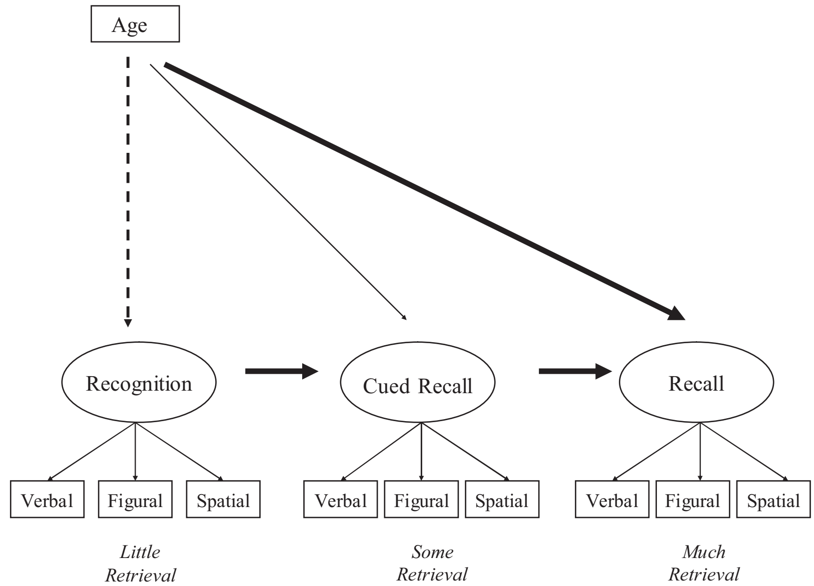

Figure 2. Mediation model that represents the theoretical relations among age and test-type constructs.

Verbal Memory

In this task, participants were presented with six lists of 25 words on the computer screen at the rate of one word per second. The words were all five and six letters long, and each list had a mean Kučera and Francis (1967) frequency rating of 77 or 78. Following each list, a recall, cued recall, or recognition memory test was administered. The task was administered on two separate sessions such that three lists were presented in Session 1 (in the following order: recall, cued recall, recognition), and three lists were presented in Session 3 in the reverse order (i.e., recognition, cued recall, recall). For each list, the participant was instructed to pay attention to the words because they would be tested on them later, but they were not informed as to which test would follow the presentation of the list so that the encoding strategies across the different lists should be equivalent.

In the recall condition, participants were instructed to recall as many words as they could remember aloud to the examiner. In the cued recall condition, participants viewed the first two letters of each word from the previous list on the computer screen and were asked to state the word aloud to the examiner. Each of the stimulus words in the cued recall task was unique with respect to the first two letters. The recognition task required participants to complete a 50-word recognition test in which they had to determine whether each presented word was old (previously presented) or new (not previously presented). The recognition test was presented on the computer screen and was self-paced by the participant. The list of

Table 1
Sample Characteristics

	Age group								
	18–39		40–59		60–94		Total		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Age *r*
N	102		117		108		327		
Age	26.30	6.40	50.70	5.70	71.30	7.80	49.90	19.20	
Female (%)	63		76		68		.05		
Health	1.80	0.80	2.10	0.80	2.10	0.80	2.00	0.80	.17**
Education (years)	14.90	2.30	15.30	2.70	16.70	3.60	15.60	3.00	.26**
MMSE	28.80	1.60	28.50	1.80	28.60	1.50	28.60	1.60	−.08
Vocabulary	13.40	3.40	11.90	3.20	13.80	2.60	13.00	3.20	.05
Digit symbol	10.50	3.00	10.80	3.00	11.50	2.50	10.90	2.80	.15**
Logical memory	11.50	2.80	11.00	3.50	12.60	2.90	11.70	3.10	.14**
Recall	12.10	3.00	12.10	3.40	12.60	3.10	12.30	3.20	.07

Note. Health is a self-reported rating on a scale of 1 (*excellent*) to 5 (*poor*). Age *r* is the correlation of each measure with age. MMSE = Mini-Mental State Examination (Folstein et al., 1975).
** $p < .01$.

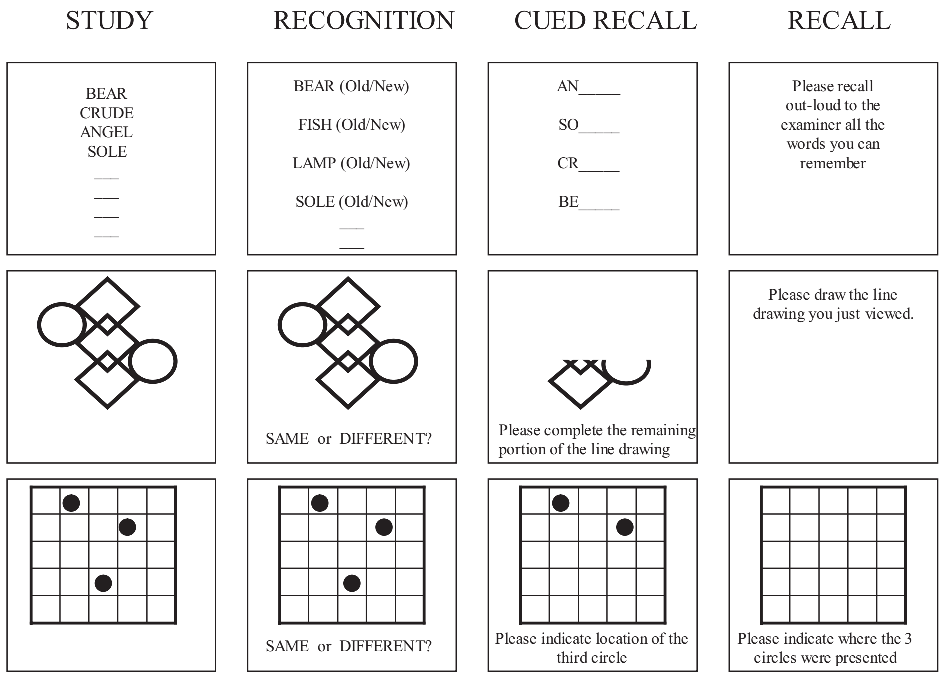

Figure 3. Examples of stimuli across material (verbal, figural, and spatial) and across test-type (recognition, cued recall, and recall).

new words in the recognition test also had a mean Kučera and Francis (1967) rating of 77 or 78. The percentages of correctly recalled words were recorded in the recall and in the cued recall tests. Recognition performance was assessed with hit rate (HR) and false-alarm rate (FAR), and a corrected recognition rate (CR) was calculated by subtracting FAR from HR.

Figural Memory

Figural memory was operationalized as memory for information based on the relation of elements to one another and was assessed with line drawing stimuli.[3] Participants were presented with abstract line drawings on the computer screen, and a memory test was administered after each stimulus.

Each drawing appeared on the computer screen for 3 s. Following the presentation of the drawing, the participants engaged in an 8-s distractor task in which they were instructed to circle as many letters *P* as possible within a set of *P*s and *Q*s. The purpose of the distractor task was to displace the stimulus material out of working memory and into episodic memory storage, given that a memory test followed each individual stimulus. A recall, cued recall, or recognition test was presented after the distractor task. As in the verbal memory task, participants were not informed as to what type of test followed each stimulus presentation, thus encoding strategies were presumably held constant.

Participants viewed a total of 54 drawings, completing 18 recall trials, 18 cued recall trials, and 18 recognition trials. An answer booklet was created that contained separate test forms for each trial. In the recall test, participants were provided with a blank page and asked to reproduce the line drawing they had just viewed. In the cued recall test, they were presented with a portion of the drawing and were asked to fill in the missing elements of the figure. The recognition test was a same–different recognition test in which the participant viewed a line drawing on a page in the answer booklet and decided whether the figure was the same (as the line drawing just presented) or different in any way (see Figure 3 for examples). The order in which test types were presented was randomized, with the qualification that no more than three tests of one kind could occur consecutively, and all participants received the same random order.

A total of 54 drawings were prepared, and the drawings were randomly assigned to the recall, cued recall, or recognition condition. Thus the stimuli in each condition did not differ systematically from one another.

Responses in the recall and the cued recall tests were scored with a three-point rating scale. A drawing was awarded three points for an entirely correct reproduction of the stimulus. Two points were awarded if the structures were mostly correct. One point was awarded if the drawing had only slight similarity to the original line drawing. Zero points were awarded if the drawing showed little or no resemblance to the original line drawing. In the recognition test, HR, FAR, and CR were recorded.

Spatial Memory

Spatial memory was operationalized as memory for locations in a spatial array (similar to the stimuli used by Ichikawa, 1983). In

[3] The drawings were constructed according to the following guidelines: (a) The figures did not resemble familiar designs; (b) the figures were all approximately the same size; (c) the minimum number of lines to create each figure was 8 (a circle was considered 1 line, whereas a square was considered 4 lines), and the maximum number of lines was 20; and, (d) the drawings were approximately similar in subjective complexity.

this task, participants were presented with 5 × 5 grids on the computer in which three of the cells within the grid were occupied by circles. The grid patterns were presented for 3 s on the computer screen. As in the figural memory task, each stimulus presentation was followed by an 8-s distractor task prior to each memory test in an attempt to displace the information to be retrieved from episodic memory storage.

Participants completed a total of 18 recall tests, 18 cued recall tests, and 18 recognition tests. An answer booklet for the spatial task was created with separate pages for each trial. In the recall test, participants were provided with a page containing a blank 5 × 5 grid and asked to mark the locations of the circles. In the cued recall test, participants were given the location of two of the circles in the grid and were asked to indicate the location of the third circle by drawing a circle in the correct cell (see Figure 3 for an example). In the recognition test, participants completed a same–different test and had to decide whether the displayed pattern was identical (same) or different from the pattern just presented. Fifty-four grid patterns were constructed and the grids were randomly assigned to the recall, cued recall, or recognition condition so the stimuli in each condition did not differ systematically from one another.

Error scores have been used in previous studies of spatial recall (Pezdek, 1983) and were recorded as a measure of accuracy. Partial credit was based on deviations from the correct positions. For example, if all three circles were in the correct location, the participant received an error score of zero. Euclidean distance of the location of the incorrect circle to the correct cell location was calculated, and participants received error scores based on the average distance of each circle from its correct location. In the situation in which no circles were in the correct location, the closest target circle was used as the reference for each incorrect circle location. HR, FAR, and CR were recorded in the recognition test.

For all of the recognition measures d' was also calculated (Z[HR] − Z[FAR]; e.g., Macmillan & Creelman, 1990), and the correlation between CR and d' was .98 in verbal recognition memory, .97 in figural recognition memory, and .98 in spatial recognition memory. Because these measures were so highly correlated, only CR was used in the subsequent analyses (see Snodgrass & Corwin, 1988, for a description of why d' and CR are equally sensitive measures of memory).

Modeling Procedure

Model fit was evaluated with multiple indices as recommended by Hu and Bentler (1998). Chi-square was selected because it is one of the most commonly used indices, and it is an indicator of how close the population covariance matrix matches the observed model covariance matrix. The critical ratio (χ^2/df; Bollen, 1989) is also reported, and because the chi-square is divided by degrees of freedom, the complexity of the model is accounted for. Another commonly used index, the root-mean-square error of approximation (RMSEA), was selected, as were Bentler's comparative fit index (CFI) and the Tucker–Lewis index (TLI) because although they are sensitive to model misspecification, they are less sensitive to distribution and sample size (for a detailed discussion of fit indices, see Hu & Bentler, 1998). In addition, because several of the models are nested, a direct test of significance can also be conducted by comparing the difference in chi-square per difference in degree of freedom across the different models.

Although different cut-off scores have been used to determine how well a model fits the data, an RMSEA of ≤.06 is generally assumed to reflect a good fit to the data (Hu & Bentler, 1999), with scores between .05 and .08 indicating acceptable fits (Browne & Cudeck, 1993). Typically a TLI or CFI value of ≥.95 represents a good fit to the data (Hu & Bentler, 1999), although values of ≥.90 have also been used to indicate a good fit.

All confirmatory factor analyses were conducted with Amos 5.0 (Arbuckle, 2003), and one loading on each factor was set to a value of 1.00 to establish a metric for estimation. The critical value for all comparisons throughout the article was $p < .01$. Given that an appropriate model of episodic memory is established, a secondary goal of this project is to determine whether the model is invariant across age. To examine the invariance of the model, age was divided by double decades to create three age groups: 18–39 years old ($N = 102$), 40–59 years old ($N = 117$), and 60 years old and over ($N = 108$). The groups are divided in this manner to create a young adult, middle adult, and older adult group that are approximately the same size.

The procedure used to evaluate age invariance involved a series of invariance assessments that become progressively more stringent (e.g., Taub, McGrew, & Witta, 2004). *Measurement invariance tests* are tests that are concerned with the relations between the manifest, measured variables and the latent construct, and *structural invariance* refers to the tests concerning the relations among the latent constructs (e.g., Byrne, Shavelson, & Muthén, 1989; Nyberg, 1994). Three tests of invariance were evaluated in the current study (configural, metric, and structural). Configural invariance (Horn, McArdle, & Mason, 1983) specifies that the factor structure remains the same across groups. Metric invariance (a form of measurement invariance), called *Invariance 1* in this article, is a more stringent test of invariance and requires that the unstandardized coefficients from latent constructs to the manifest variables are the same across groups (Horn & McArdle, 1992). The third test of invariance, called *Invariance 2*, is a test of structural invariance and constrains both the factor loadings and the interfactor covariances to be the same across groups.

The mediation model portrayed in Figure 2 was used to examine whether there were unique age-related effects on cued recall after first considering the age-related effects on recognition and whether there were unique age-related effects on recall after partialling out the age-related effects on cued recall.

Results

The Results section is organized into four parts. The first part reports the results from the individual tasks designed to assess potentially distinct aspects of memory. The second section describes the results from the series of model comparisons designed to identify the most meaningful representation of episodic memory with these tasks and includes an assessment of construct validity. The third section reports the results of age invariance analyses on the model selected as the most meaningful representation of episodic memory. The final section reports the results from analyses designed to assess the unique age-related effects on the different test-type constructs by examining the age mediation models.

Individual Tasks

The estimated reliability, mean, standard deviation, and age correlation of each individual variable is reported in Table 2. Reliability was assessed at the item level with Cronbach's alpha for all of the measures. For recognition measures, reliability was based on percentage correct because the HR–FAR is not applicable for individual items. All but two of the variables had relatively high reliability, with estimated coefficient alphas ranging from .73 to .93.

The age-squared terms, representing the quadratic relations of the variables to age after controlling for linear age relations, were not significantly different from zero for any of the variables, and thus only linear age relations were included in subsequent analyses.

To portray the linear age trends, each variable was converted into a *z* score and plotted as a function of age by decade in Figure 4. Spatial memory performance is measured by error scores, so the relationship between age and the unconverted scores is positive. In an effort to ease interpretations of the age trends, the signs of the spatial memory recall and cued recall scores were reversed for all subsequent figures. Inspection of Figure 4 shows that age appears to be monotonically related to recall and cued recall performance across the different modality types. Recognition performance, especially verbal recognition, appears to have less of a relationship with age.

The correlations among the nine variables are reported in Table 3, with the reliabilities reported in the diagonal. Inspection of the correlations indicates that all the correlations among the variables are significant at the .01 level, and the variables within modality domains (hypothesized to represent verbal, figural, and spatial constructs) are more highly correlated with one another than are variables between modality domains. Specifically, the mean correlation of the variables within the verbal domain is .35, and it is .26 between the verbal variables and the nonverbal variables. The mean correlation of the variables within the figural domain is .70, but it is only .44 between the figural variables and the nonfigural variables, and the mean correlation of the variables within the spatial domain is .64, but it is only .41 between the spatial variables and the nonspatial variables.

Table 2
Reliabilities, Means, Standard Deviations, and Linear Age Correlations for the Nine Manifest Variables

Variable	Reliability	*M*	*SD*	Age *r*
Verbal memory				
Recall	0.74	4.07	2.35	$-.40^{**}$
Cued recall	0.73	7.66	2.56	$-.27^{**}$
Recognition	0.88	0.48	0.24	$-.06$
Figural memory				
Recall	0.92	0.93	0.42	$-.55^{**}$
Cued recall	0.93	1.76	0.56	$-.57^{**}$
Recognition	0.56	0.50	0.22	$-.15^{**}$
Spatial memory				
Recall	0.84	0.73	0.33	$.40^{**}$
Cued recall	0.75	0.60	0.37	$.43^{**}$
Recognition	0.45	0.45	0.26	$-.35^{**}$

Note. The mean recognition measures are the corrected recognition rates (hit rate−false alarm rate) for each modality type. Age *r* is the correlation of each measure with age. In verbal memory, the recall and cued recall measures correspond to the total number of words correctly recalled from a list of 25. For figural memory, the recall and cued recall variables are the mean subjective ratings of line drawings, ranging from 0 to 3. The recall and cued recall variables in the spatial memory task are error scores representing the distance from the correct positions.
$^{**}p < .01$.

Verbal Memory

Verbal memory was measured with recall, cued recall, and recognition tasks. As reported in Table 2, both the recall and cued recall measures were negatively correlated with age ($r = -.40$ and $r = -.27$, respectively). However, verbal recognition performance was not significantly correlated with age. The mean number of words recalled (out of 25) was just 4.07, which is lower than might be expected. For instance, the mean number of words recalled from the first list of the Word Recall subtest for the WMS–III for these same participants was 6.38. The somewhat low performance on the verbal recall task could be a function of the difficulty of the task, namely, the quick presentation rate and the large number of words.

Figural Memory

Recall and cued recall performance of the abstract line drawings was assessed with a rating scale ranging from zero to three by two independent raters. The mean interrater reliability for the recall measures was .89, and for the cued recall measures it was .87.[4] Another way to examine the consistency among raters is to create a difference score (i.e., Rater A score − Rater B score) and to examine the percentage of scores that are greater than some criterion. Because the difference scores could range from −3 to +3, one can argue that difference scores of greater than ±1 would indicate that the two raters disagreed. For the recall task, 98.4 % of the difference scores were between 0 and ±1, indicating strong agreement on ratings. Just 1.5% of the scores were in the ±2 range, and .07% of the difference scores were within the ±3 range. The findings for the cued recall ratings were similar in that 98% of the differences scores were not greater than (the absolute value of) 1, 1.8% of the scores were in the ±2 range, and .12% in the ±3 range.

All three measures of figural memory were negatively correlated with age, although the relation of recognition performance to age was not as great as with the other measures. The average score for participants in the recall task was .98 (out of 3.00), and similar to the verbal task, participants performed almost twice as well in the cued condition ($M = 1.76$).

Although participants were instructed that the distractor task was as important as the memory task, it is possible that they divided their efforts unequally and performed less well on the distractor task in order to perform better on the memory task. The

[4] To ensure the validity of the subjective rating scale, detailed scoring criteria for six randomly chosen line drawings were created. Scores based on the detailed criteria were subsequently compared for a subset of 42 participants, with the scores based on the subjective criteria. The correlation of the scores from the detailed scoring criteria for the two independent raters was .86, and the correlation between the ratings and the scores based on the criteria was .87.

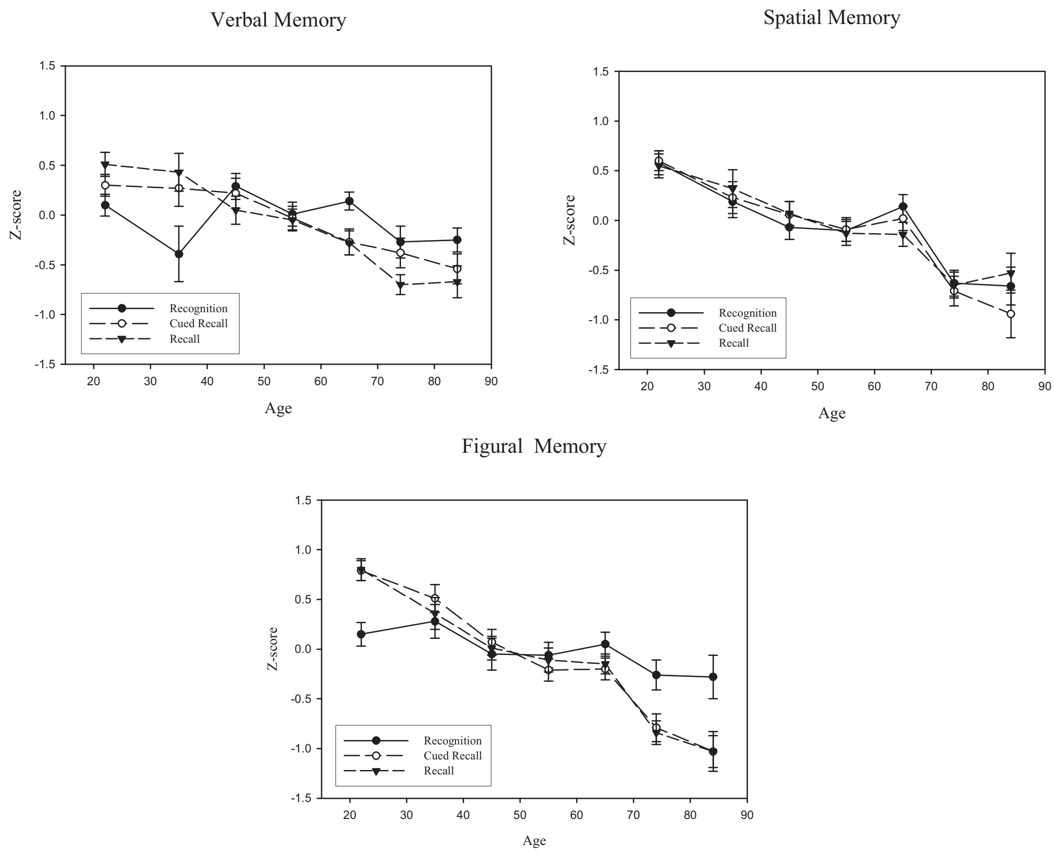

Figure 4. Means and standard errors by decade of performance on the recognition, cued recall, and recall for verbal memory, spatial memory, and figural memory.

average number of letters circled during the 8-s delay was used as the indicator of distractor task performance. The partial correlation, after controlling for age, between figural memory performance and distractor task performance was calculated for each test type. The partial correlation for recognition and distractor task performance during the recognition tests was .22 ($p < .01$), for cued recall and distractor task performance during cued recall it was .27 ($p < .01$), and for recall and distractor task performance during recall it was .20 ($p < .01$). These findings suggest that participants were not performing less well in the distractor task in order to preserve figural memory performance. In fact, higher performance on the distractor task was associated with better figural memory performance.

Spatial Memory

It is important to note that spatial memory performance was measured by the degree of error in assigning a location to a circle within the grid. Thus the positive correlations with age indicate that increased age is associated with an increase in the magnitude of errors. Recognition performance, however, is measured as in both the verbal memory and figural memory tasks and is the corrected recognition rate (i.e., HR − FAR), and so its relationship with age is negative.

The partial correlation, after controlling for age, between distractor task performance and spatial memory performance was .22 ($p < .01$) for recognition, −.20 ($p < .01$) for cued recall, and −.13 (*ns*) for recall. Because cued recall and recall performance were measured with error scores, a negative partial correlation indicates that better performance on the distractor task was associated with fewer errors on the cued recall and recall tasks.

Structural Models

Two structural organizations of episodic memory were examined by using the same nine variables. The material modality models examined the relations among the hypothesized verbal, figural, and spatial factors; and the test-type models examined the relations among the hypothesized recall, cued recall, and recognition factors.

Modality of Material Models

The fits of three nested models were examined to determine which model was the best representation of episodic memory in terms of modality of material. Because the models were nested, a direct comparison can be made across the three models. The baseline model was a one-factor model in which all nine manifest variables were subsumed under a single episodic memory construct. The two-factor model consisted of a verbal memory factor and a combined figural–spatial factor. The three-factor model was

Table 3
Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9
1. Verbal recall	(.74)								
2. Verbal cued	.45	(.73)							
3. Verbal recognition	.27	.29	(.88)						
4. Figural recall	.45	.31	.15	(.92)					
5. Figural cued	.46	.35	.15	.86	(.93)				
6. Figural recognition	.29	.28	.17	.59	.64	(.56)			
7. Spatial recall	.33	.22	.14	.65	.66	.53	(.84)		
8. Spatial cued	.30	.27	.16	.68	.71	.55	.73	(.75)	
9. Spatial recognition	.27	.20	.14	.52	.52	.46	.61	.58	(.45)

Note. Reliabilities are reported in parentheses. All *ps* $< .01$.

comprised of separate verbal memory, figural memory, and spatial memory factors (as depicted in the top portion of Figure 1).

The results are presented in the top portion of Table 4. Inspection of the table reveals that the one-factor model provided the poorest fit to the data, $\chi^2(27) = 185$, RMSEA $= .13$, followed by the two-factor model, which was better than the one-factor model but was still a poor fit overall, $\chi^2(26) = 124$, RMSEA $= .11$. The three-factor model clearly provided the best fit to the data, $\chi^2(24) = 34$, RMSEA $= .04$. Both the CFI and the TLI were .99, indicating that the fit was very good. Because the models are nested, a direct test of significance can also be conducted by comparing the difference in chi-square with the differences in degrees of freedom. As predicted, the two-factor model fit significantly better than the one-factor model, and the three-factor model fit significantly better than both the one-factor and the two-factor models at the $p < .01$ level. ($N = 327$, for all chi-squares.)

Construct validity of a model is evaluated by examining convergent and discriminant validity of the latent constructs. Convergent validity is demonstrated when the variables hypothesized to represent the same construct have significant variance in common, as manifested in moderate to large standardized coefficients from the construct to the variables. The three-factor modality material model, with standardized coefficients and construct correlations, is depicted in Figure 5. The three constructs of the three-factor model can be considered to exhibit moderate to strong convergent validity because all of the lower order loadings are significantly different from zero and are in the moderate to large range (i.e., eight of the nine standardized coefficients have absolute values of .60 or greater). This suggests that verbal memory, figural memory, and spatial memory constructs do exhibit convergent validity. Discriminant validity is demonstrated when variables hypothesized to represent different dimensions of individual differences are distinct from one another. In this case, discriminant validity can be measured in two ways. First, inspection of the correlations among the latent variables indicates that the verbal memory construct is only moderately correlated with the figural memory and spatial memory constructs, and although the correlation between the figural memory and spatial memory constructs is substantial ($r = .85$), the 99% confidence interval does not include 1.00 (.79–.90). Discriminant validity is also demonstrated because the three-factor model fits significantly better then the one- and two-factor models, providing further evidence of distinct memory factors.

The correlations among the modality of material constructs and age were calculated, and, as expected, each of the constructs was significantly correlated with age. Age was correlated with the verbal memory construct (−.48), with the figural memory construct (−.60), and with the spatial memory construct (−.49).

Test-Type Models

Four nested test-type models were compared: a one-factor model, 2 two-factor models (2A and 2B), and a three-factor model. Given that the same nine variables are used, the one-factor test-type model is identical to the one-factor material modality model. The 1st two-factor model (2A) is comprised of a recognition factor and a second factor that combines the cued recall and recall

Table 4
Fit Statistics for the Competing Models

	Goodness-of-fit indices					
Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA
1-factor	185.48	27	6.87	0.90	0.86	0.13
Modality of material models						
2-factor (Verbal/Figural & Spatial)	124.17	26	4.78	0.94	0.91	0.11
3-factor (Verbal, Figural, Spatial)	33.87	24	1.41	0.99	0.99	0.04
Test-type models						
2A: 2-factor (Recognition/Cued & Recall)	183.00	26	7.04	0.90	0.86	0.14
2B: 2-factor (Recall/Cued & Recognition)	Inadmissible solution					
3-factor (Recall, Cued, Recognition)	Inadmissible solution					

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.

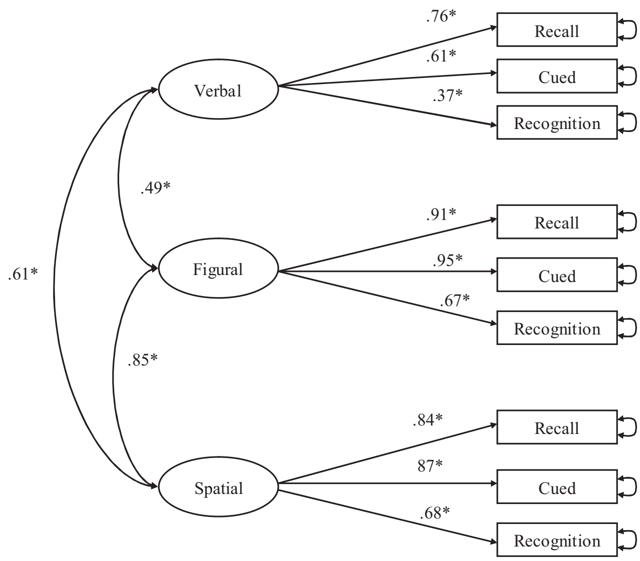

Figure 5. Three-factor material of modality model with standardized coefficients.

variables. As is evident in Table 4, the fit of this model was comparable to, or slightly worse than, the one-factor model. The 2nd two-factor model (2B) consisted of a recall factor that was distinct from a factor comprised of cued recall and recognition variables. In this model, the resulting model fit was not admissible because the covariance matrix was not positive definite (i.e., the correlation estimates between the two factors were so large that the factors could not be distinguished). The differences between the 2 two-factor models are that in Model 2A the cued recall variables loaded on a factor with the recall variables, whereas in the Model 2B the cued recall variables loaded on a factor with the recognition variables. The solution for the final model, a three-factor model distinguishing recall, cued recall, and recognition factors, was also inadmissible because of the covariance matrix not being positive definite.

The results of the model comparisons indicate that the correlation estimates among the constructs of recall, cued recall, and recognition are so large that the constructs cannot be distinguished from one another, at least when the verbal, figural, and spatial material are considered together.

Given that the findings of the test-type model fit comparisons indicated that verbal memory, figural memory, and spatial memory were distinct constructs, several additional test-type models were compared post hoc. Namely, each test-type model was reanalyzed with the verbal, figural, and spatial memory manifest variables examined separately. Because the memory tasks were administered across two sessions, there were, for example, two sets of verbal recognition tasks, two sets of verbal cued recall tasks, and two sets of verbal recall tasks (i.e., one set of each administered during Session 1 and during Session 3). In these new models, each factor was created with two manifest variables derived from the score on each session. For example, the verbal recall factor was comprised of a verbal recall variable from Set 1 and a verbal recall variable from Set 2.

Fit statistics and correlations of the factors for the new models are presented in Table 5. Inspection of the fit statistics indicates that both the two-factor model (i.e., 2B), which distinguished a recall factor from a combined cued recall and recognition factor, and the three-factor model provided good fit to the data for the verbal memory, with the three-factor model providing a slightly, but not significantly, better fit. In a direct comparison of change in chi-square per degree of freedom change in the four nested models, the three-factor model fit significantly better than both the one-factor model and the first two-factor model at the $p < .01$ level.

Within the figural domain, both the one-factor and the two-factor model that distinguished recognition from a combined cued recall–recall factor provided reasonably acceptable fits to the data. The one-factor fit better as indicated by comparison of the TLI and RMSEA statistics, although the difference in change in chi-square per degree of freedom change was not significant. However, the correlation estimate between the two factors was 1.03 (with 99% confidence intervals including 1.00; 0.83–1.13), providing additional evidence that the one-factor model best described the relations among the test-type variables within the figural domain because the constructs in the two-factor model clearly lack discriminant validity.

Table 5 also shows that only the one-factor test-type model was admissible within the spatial domain. The interfactor correlations for the three-factor models for each modality of material are reported in Table 6, which clearly shows that many of the correlation estimates among the factors within the figural and spatial domain are 1.00, or not different from 1.00, once again demonstrating a lack of discriminant validity between the hypothesized factors or constructs. It can be seen that within the verbal domain, the three factors can be distinguished from another, with correlations ranging from .60 to .73.

Age Invariance Models

After determining that the material modality three-factor model was the best representation of episodic memory in the cross-sectional sample (as determined by its superior fit and evidence of construct validity), the structural invariance of the model was examined across three age groups. Configural invariance was first evaluated by specifying the structure of the model to be the same across the three groups. Nearly all loadings were significant at the $p < .01$ level for all age groups, and the fit of the configural model was good, $\chi^2(72) = 91.27$, RMSEA $= .03$, suggesting that the three-factor model was most likely an appropriate representation of memory across each group.

Metric invariance was examined by constraining the factor loadings to be equivalent across the age groups (Invariance 1) and by comparing the fit to the configural invariance model. For a model to exhibit metric invariance the fit of the Invariance 1 model should not be significantly different from the fit of the configural model. That is, the fit of the model should not be significantly worse after constraining the unstandardized loadings to be equivalent across the age groups, indicating that the values of the loadings cannot be detected to differ across groups. As can be seen in Table 7, the change in chi-square per degree of freedom change between the configural model and the Invariance 1 model was not significant.

In addition, the fit of the more restrictive test of invariance (Invariance 2) was also not significantly worse, as compared with the baseline configural model. In this case, both the factor loadings

Table 5
Fit Statistics and Intercorrelations Among the Constructs in the Separate Test-Type Models

Model	Goodness-of-fit indices					
	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA
Verbal						
1-factor	33.29	9	3.70	0.92	0.80	0.09
2A: 2-factor (Recognition/Cued & Recall)	28.14	8	3.52	0.93	0.81	0.09
2B: 2-factor (Recall/Recognition & Cued)	19.31	8	2.41	0.96	0.90	0.07
3-factor	14.34	6	2.39	0.97	0.90	0.07
Figural						
1-factor	52.86	9	5.87	0.97	0.93	0.12
2A: 2-factor (Recognition/Cued & Recall)	52.76	8	6.60	0.97	0.97	0.13
2B: 2-factor (Recall/Recognition & Cued)	47.75	8	Inadmissable solution			
3-factor	47.46	6	Inadmissable solution			
Spatial						
1-factor	17.72	9	1.97	0.99	0.97	0.06
2A: 2-factor (Recognition/Cued & Recall)	16.31	8	Inadmissable solution			
2B: 2-factor (Recall/Recognition & Cued)	17.68	8	Inadmissable solution			
3-factor	15.98	6	Inadmissable solution			

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.

and the factor covariances were constrained to be the same across the age groups. This finding suggests that the magnitude of the correlations among the constructs is not significantly different across the age groups. Table 8 reports the correlations, with 99% confidence intervals, among the constructs for each age group within the configural model. Inspection of the table indicates that although the correlation between the verbal memory construct and the spatial memory construct varied across the age groups ($r = .40$ for the young group, .14 for the middle group, and .60 for the older group, respectively), they were not significantly different as demonstrated by the observation that there was some overlap among confidence intervals of the correlations. It is possible that there are differences in the magnitude of the correlations, but power is too low to detect these differences.

The original test-type models had inadmissible solutions because the estimates for the correlations among the constructs were so large. Consequently, several post hoc models were compared in which test-type models were examined separately for the verbal, figural, and spatial information. Because the three-factor model distinguishing between recognition, cued recall, and recall provided the best fit for the verbal memory tasks, and the one-factor model fit best for figural and spatial memory tasks, these models were examined for age invariance as well. Given that the figural and spatial memory models were one-factor models, only configural and Invariance 1 could be examined (because the Invariance 2 model requires constraining interfactor covariances). Table 9 presents the results from the age invariance analyses, in which it can be seen that the three-factor verbal memory model exhibited configural, metric, and structural invariance. Correlations and 99% confidence intervals for the three-factor model distinguishing among the constructs of recognition, cued recall, and recall within the verbal domain are presented in Table 10. The correlations across the age groups appear to be moderately similar, although the confidence intervals are fairly large, indicating a lack of precision in detecting differences.

The one-factor spatial memory model also exhibited configural and metric invariance. Although the one-factor figural memory model did exhibit evidence of configural invariance, $\chi^2(27) = 64.7$, RMSEA = .07, the Invariance 1 model fit significantly worse than did the baseline model, indicating that the relations of the variables to the construct may not be the same across the age groups. Because the figural model did not exhibit metric invariance when all the loadings were constrained, a possible next step in invariance testing is to examine whether the model exhibits partial invariance (Byrne et al., 1989), which refers to cases in which only some model parameters are invariant. To determine whether there was one specific loading that was driving the significantly worse fit of the Invariance 1 model, each loading was allowed to vary across groups one at a time. When the loading from Set 2 recognition to the general figural memory factor was

Table 6
Interfactor Correlations and 99% Confidence Intervals (in Parentheses) of the Test-Type Models for the Three-Factor Models

Model	Verbal	Figural[a]	Spatial[a]
Recall/Cued	0.73** (0.56–0.90)	1.03** (1.00–1.06)	0.99** (0.91–1.07)
Cued/Recognition	0.65** (0.36–0.94)	0.98** (0.83–1.13)	1.08** (0.84–1.31)
Recognition/Recall	0.60** (0.35–0.86)	0.97** (0.82–1.12)	1.11** (0.87–1.35)

[a] Covariance matrix is not positive definite.
** $p < .01$.

Table 7
Goodness-of-Fit Indices for the Age Invariance Models for the Three-Factor Modality of Material Model

Model	Goodness-of-fit indices								
	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA	$\Delta\chi^2$	Δdf	$p < .01$
Configural	91.27	72	1.27	0.99	0.98	0.03	Baseline model		
Invariance 1	113.91	84	1.36	0.98	0.97	0.03	22.64	12	No
Invariance 2	125.50	90	1.39	0.97	0.97	0.04	34.23	18	No

Note. CFI = comparative fit index; TLI = Tucker–Lewis Index; RMSEA = root-mean-square error of approximation.

allowed to vary across groups, the Invariance 1 model no longer fit significantly worse than did the configural model, suggesting that it was this one loading that was significantly different across the groups. Specifically, the standardized factor loadings in the configural model were .51 for the young group, .41 for the middle group, and .44 for the older group. The lack of systematic difference across the age groups in the magnitude of the loadings and the absence of an a priori hypothesis to explain the fluctuation across the groups suggest that this result is likely spurious. It is important to replicate this finding before generating post hoc explanations.

Age Mediation Models

The original intention was to examine the mediation model with verbal, figural, and spatial material across the three test-type constructs (as depicted in Figure 2). However, as reported in Table 4, the best-fitting test-type model was the one-factor model. Examining the mediation model across the three constructs would not be meaningful because the correlation estimates among the recognition, cued recall, and recall constructs were so large that the constructs could not be distinguished from one another. Thus it was only meaningful to examine the mediation model, as depicted in Figure 2, with verbal material since it was only within the verbal domain the test-type constructs were distinct. As predicted, there were no statistically significant, age-related effects on recognition (−.14, *ns*), but there were unique age-related effects on both cued recall (−.27, $p < .01$) and recall (−.27, $p < .01$).

Discussion

There were three goals of the current project. The primary goal was to determine the most meaningful representation of episodic memory in terms of verbal, figural, and spatial material, and of tests of recognition, cued recall, and recall. A series of theory-driven model comparisons indicate that the three-factor modality of material model that distinguished among verbal memory, figural memory, and spatial memory was the most meaningful representation of episodic memory. The three-factor model exhibited both convergent and discriminant validity such that the variables hypothesized to represent the construct of interest were correlated with one another, and the correlations among the constructs hypothesized to represent different constructs were significantly less than 1.00.

Table 8
Interfactor Correlations and 99% Confidence Intervals (in Parentheses) Among the Material of Modality Factors Across Age Groups in the Three-Factor Configural Model

Factor	Young	Middle	Old
Verbal/Spatial	.40** (.11–.68)	.14 (−.10–.38)	.60** (.35–.84)
Verbal/Figural	.39** (.11–.67)	.52** (.29–74)	.59** (.37–.81)
Spatial/Figural	.86** (.76–.96)	.81** (.70–.92)	.80** (.66–.93)

** $p < .01$.

This finding is consistent with past research across the cognitive univariate, correlational, and neurobiological approaches, which have provided evidence that visual, figural, and spatial memory represent separate constructs (e.g., Herrmann et al., 2002; Klauer & Zhao, 2004; Wechsler, 1997a; Wilde et al., 2003). Although the three constructs were distinct, the figural memory and spatial memory constructs were highly correlated ($r = .85$), suggesting that individuals who perform well on tests of memory for locations (i.e., spatial memory) also perform well on tests of memory for line drawings (i.e., figural memory), and vice versa. This finding is also consistent with past research because figural memory and spatial memory have often been considered collectively (e.g., Jenkins, Myerson, Joerding, & Hale, 2000). However, it may also be the case that the figural and spatial constructs are more highly correlated with one another than with the verbal memory construct because of the nature of the stimuli and the manner in which they were tested was more similar across the figural and spatial tasks than in the verbal task.

The model comparisons also indicated that although recall, cued recall, and recognition could be viewed as distinct constructs within the verbal domain, this was not the case within the figural and spatial domains. Within the figural and spatial domains, the estimated interfactor correlations among the hypothesized constructs corresponding to recognition, cued recall, and recall constructs were so large that separate constructs could not be distinguished. This finding suggests that the pattern of retrieval for figural and spatial information may be qualitatively different as compared with verbal material. Namely, how information is retrieved may be irrelevant for individual differences in figural and spatial memory but may be important as a determinant of individual differences in verbal memory.

What can account for these differences? What makes verbal memory different from figural and spatial memory? By their nature, figural and spatial stimuli are abstract, designed specifically to prevent verbal encoding of the information, and tend to be more complex because they are almost entirely composed of nonsense forms or patterns, which may prevent individuals from creating some type of "organizational structure" that may help in

Table 9
Goodness-of-Fit Indices for the Age Invariance Models for Test-Type Models for Verbal, Figural, and Spatial Memory

	Goodness-of-fit indices								
Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA	$\Delta\chi^2$	Δdf	$p < .01$
				Verbal memory					
Configural	26.78	18	1.49	0.96	0.87	0.04		Baseline model	
Invariance 1	33.03	24	1.38	0.96	0.90	0.03	6.25	6	No
Invariance 2	39.54	30	1.32	0.96	0.92	0.03	12.76	12	No
				Figural memory					
Configural	64.66	27	2.40	0.97	0.92	0.07		Baseline model	
Invariance 1	88.65	37	2.40	0.95	0.92	0.07	23.99	10	Yes ($p = .008$)
				Spatial memory					
Configural	28.52	27	1.06	1.00	0.99	0.01		Baseline model	
Invariance 1	43.85	37	1.19	0.99	0.98	0.02	15.32	10	No

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.

retrieving meaningful shapes (Ameli, Courchesne, Lincoln, Kaufman, & Grillon, 1988) or meaningful words. Because there is no meaningful organizational structure to aid in the retrieval of information, the processing of figural and spatial information may be more holistic and lend itself to all-or-none processing. This all-or-none processing could account for the differences between verbal and nonverbal retrieval. The lack of distinct constructs in the figural and spatial domains among the test-type constructs suggests that there are no differences among the ability to recognize or recall (either with or without a cue) a line pattern or locations of circles within a grid. That is, the ranking of an individual, relative to the other individuals in the sample, is the same across different conditions of retrieval for the figural and spatial memory tasks (although this does not necessarily mean that performance is the same across the different conditions).

Another explanation for these differences is that familiarity may be used to assist performance in memory tests with verbal material but not with figural and spatial material. There has been considerable research examining the processes of familiarity and recollection involved in word recognition (see Yonelinas, 2002, for a review). When words are presented to an individual, a semantic network is presumably activated, and as a result a word may appear familiar although there is no clear memory of encoding the word. However, there is no comparable network to be activated in regard to figural and spatial material, and thus it is unlikely that familiarity-based processing may be used to retrieve figural and spatial information. Consequently, it is possible that only recollective processes are used in the retrieval of figural and spatial information, and because the same process is used in recognition, cued recall, and recall tasks, the test-type constructs cannot be distinguished from one another. This interpretation further suggests that the distinct test-type constructs within the verbal domain may be the result of different processes used across the different retrieval modes. Specifically, as it has been argued in the literature, familiarity and only a little retrieval processing is used in tests of recognition, whereas cued recall likely requires a combination of both familiarity and retrieval, and recall requires mainly retrieval processing.

Still another explanation for the results presented here may be the methodological differences in how the verbal memory and the nonverbal memory were assessed. Figural memory and spatial memory were tested after each stimulus presentation, whereas verbal memory was assessed after a presentation of a list of items. In addition, there was an 8-s-filled delay in which participants performed a visual–spatial distractor test in the figural and spatial memory tasks but not in the verbal memory task. These methodological differences may have affected the processing of the stimuli.

Another possible explanation for the qualitative difference in how figural and spatial information is retrieved compared with that of verbal information may be related to the fact that the delay between presentation and test was brief. That is, it is possible that

Table 10
Interfactor Correlations and 99% Confidence Intervals (in Parentheses) for the Three-Factor Test-Type Configural Invariance Model for Verbal Material

Factor	Young	Middle	Old
Recall/Cued	0.75** (0.36–1.13)	0.62** (0.33–0.91)	0.76** (0.40–1.11)
Cued/Recognition	0.29 (−0.22–0.80)	0.79** (0.31–1.27)	0.84** (0.35–1.32)
Recognition/Recall	0.62** (0.06–1.18)	0.83** (0.32–1.34)	0.47** (0.12–0.83)

** $p < .01$.

the figural and spatial information was still present in working memory, whereas most of the verbal information was no longer in working memory because of the longer list of items. As a consequence, the information did not need to be actively retrieved because it was already present, which could explain why there were no distinctions among the constructs of recognition, cued recall, and recall. Although this is a possible explanation, it seems unlikely because participants completed a visual–spatial distractor task that should have prevented them from rehearsing the information. In addition, if no retrieval was necessary, then we would expect performance on the memory tasks to be quite high and even close to ceiling performance, especially on the recognition measures. However, this was not the case because the mean hit rates for the figural and spatial memory were 0.82 ($SD = 0.17$) and 0.83 ($SD = 0.14$; a score of 1.00 would indicate perfect performance), indicating that although high, the performance was not at ceiling level. Furthermore, the mean recall score for figural memory was .93, and the mean cued recall score for figural memory was 1.76, suggesting that the task was rather difficult (a score of 3.00 indicates a perfect score). This same pattern holds true within the spatial memory domain. Because cued recall and recall performance was measured with error scores, a score of zero indicates a perfect score. The mean recall score was .73, and the mean cued recall was .60. These results clearly show the figural and spatial tasks were not particularly easy, which suggests that rather than simply maintaining the information, participants were actively retrieving the information.

Although further research needs to be conducted to eliminate the alternative explanations before conclusively arguing that it is the nature of figural and spatial material that is driving these differences in retrieval, there is additional evidence to support the idea that findings within the verbal domain do not always generalize to figural and spatial domains. For instance, Miyake and colleagues (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001) have reported that the distinction between working memory and short-term memory constructs that is evident within the verbal domain is not evident within the visual–spatial domain. In addition, Shah and Miyake (1996) reported that contrary to what they would have predicted on the basis of findings in the verbal memory domain, short-term memory and working memory performance equally predicted performance on tests of spatial ability.

The second goal of the study was to determine whether the best-fitting model (i.e., the three-factor modality of material model) was invariant across age. This model exhibited configural invariance, as well as metric and structural invariance, because the magnitude of the relations among the variables to the constructs and the interfactor correlations was the same across age. This is important because it indicates that the structure of the three-factor model is the same across age (i.e., configural invariance) and that the variables are measuring the construct to the same degree across the age groups (i.e., Invariance 1). That is, because a latent construct represents what is common among the manifest variables, this finding indicates that this representation is not different across age groups. In addition, the magnitude of the correlations among the constructs was not significantly different across the age groups, which is consistent with a review by Zelinski and Lewis (2003) in which little evidence for the dedifferentiation hypothesis in the context of latent variables was found. Although these findings appear to substantiate the assumption that age differences in performance on memory tasks are likely quantitative, it is possible that the invariance analyses lacked power to detect differences among the groups. As noted in the results section, the 99% confidence intervals are fairly large around the correlation coefficients, indicating a lack of precision. In addition, the 99% confidence intervals around the factor loading were also fairly large. Thus, it is important to replicate these age invariance findings (preferably in a larger sample to increase the precision of the estimates) before reaching strong conclusions regarding age invariance across the groups

Post hoc analyses revealed that a three-factor test-type model fit best for the verbal material, whereas a one-factor undifferentiated test-type model fit best for the figural and spatial material; these models, for the most part, also exhibited age invariance. Once again, these age invariance findings should also be replicated in another, preferably larger, sample before firm conclusions can be made.

The third goal of the study was to examine whether there are unique age-related effects on measures of cued recall and recall. The mediation model designed to examine this could only be analyzed with verbal material because the test-type constructs were not distinct within the figural and spatial domain. The results revealed that there were unique and statistically significant age-related effects on cued recall after first considering the age-related effects on recognition and unique and statistically significant age-related effects on recall after partialling out the age-related effect on cued recall.

This result is consistent with findings that increased age is associated with increased difficulty in retrieving information (Craik et al., 1996; Craik & McDowd, 1987; Whiting & Smith, 1997). An additional explanation is therefore needed for the age-related effects on cued recall and recall of verbal material that accounts for the age-related effects that are above and beyond the age-related effects on recognition. Craik and colleagues (Craik et al., 1996; Craik & Jennings, 1992; Craik & McDowd, 1987) have argued that as a result of progressively less environmental support across measures of recognition, cued recall, and recall there is an increased need for self-initiated processing across the three test types. They have argued that increased age is associated with smaller pools of processing resources, which in turn makes the self-initiated processing inherent in recall tasks more difficult with increased age.

However, an alternative explanation is that poor encoding is the root cause of poor retrieval. If information is not encoded properly, a recognition test would naturally be easier than a recall test because alternatives are presented in recognition tests, and familiarity-based processing can be used. One technique that has been used in the past to identify the cause of retrieval difficulties with increased age is to require participants to perform a secondary task during encoding or retrieval.

For example, Craik and McDowd (1987) reported that older adults performed disproportionately worse on measures of cued recall than on measures of recognition and that older adults' reaction time latencies in the secondary task were considerably higher than those of younger adults in both the cued recall and recognition tests. In addition, the differences in the latencies during the cued recall test were disproportionately higher for the older adults, suggesting that the cued recall test was occupying more resources in the older adults than in the younger adults. Whiting and Smith (1997) further examined this phenomenon by manipulating the difficulty of the secondary task during recognition and recall tasks, and they reported that as the secondary task difficulty increased, older adults, relative to younger adults, performed worse on the secondary task and that these age differences were greater in the tests of recall. However, apparently contradictory findings were reported by Park et al. (1989), who in two experiments presented a secondary task during encoding, retrieval, or both. They report an age interaction with memory performance such that older adults' memory performance declined more than that of younger adults when a secondary task was presented during encoding but not during retrieval. In fact, many studies have reported that divided attention at encoding adversely affects memory performance more so than divided attention at retrieval (N. D. Anderson, Craik, & Naveh-Benjamin, 1998; Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik et al., 1996).

Each of these studies used different stimuli and different secondary tasks, and thus it is possible that the inconsistent results are a consequence of the diverse paradigms. The use of latent variables in the current study eliminates some of the task-specific influences that may be operating in prior studies, and also creates variables that are, in theory, perfectly reliable. The results of this study nicely demonstrate that there is something special about the aging of recall and cued recall (i.e., there are unique and statistically independent age-related effects on cued recall after partialling out the age-related effect on recognition, and there are unique and statistically independent age-related effects on recall after partialling out the age-related effects on cued recall). Although this study was not designed to determine the source of the age-related differences in memory test type, the decreased processing resource hypothesis can be indirectly tested by examining the relations of the memory test types with constructs that may be considered resource variables. Because participants in this study also completed multiple tests of vocabulary, processing speed, and fluid ability, it is possible to examine, in a context of structural equation models, whether there are differential effects of each of these core cognitive abilities on verbal recognition, cued recall, and recall. In a model in which the three verbal memory constructs were correlated and in which processing speed, vocabulary, or fluid ability acted as a latent predictor, there were no differences in the magnitude of the predictors across the memory type constructs. Specifically, fluid ability and processing speed were significant predictors of all three of the verbal memory constructs at the $p < .01$ level, and vocabulary was not a significant predictor on any of the verbal memory constructs. The results of these exploratory analyses are therefore inconclusive—perhaps because none of the constructs adequately measured "pools of processing resources," which may be better operationalized as working memory capacity, for example.

In conclusion, the results of this study indicate that figural and spatial material may be retrieved differently than verbal material. This finding implies that test type is important within the verbal domain but that test type may be irrelevant in figural and spatial tasks, at least in terms of individual differences. This also suggests that when developing memory theories, different types of material should be included in the research because the relations to other variables may differ as a function of the material type. In addition, this study provides empirical evidence to support the assumption that differences in memory performance across age groups are quantitative rather than qualitative. Finally, it provides additional support for the idea that there are age-related effects on cued recall that are unique and independent of the age-related effects on recognition and that there are also age-related effects on recall that are unique and independent of the age-related effects on cued recall.

References

Ameli, R., Courchesne, E., Lincoln, A., Kaufman, A. S., & Grillon, C. (1988). Visual memory processes in high-functioning individuals with autism. *Journal of Autism and Developmental Disorders, 18,* 601–615.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.). *Memory* (2nd ed., pp. 237–313). San Diego: Academic Press.

Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: Evidence from divided attention costs. *Psychology and Aging, 13,* 405–423.

Arbuckle, J. L. (2003). Amos 5.0 [Computer software]. Chicago, IL: SPSS.

Atkinson, R. C., Hertmann, D. J., & Wescourt, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 101–146). Potomac, MD: Erlbaum.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Fourth International Symposium on Attention and Performance* (pp. 583–611). New York: Academic Press.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113,* 518–540.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 445–455). Newbury Park, CA: Sage.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30,* 542–579.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Cabeza, R., Kapur, S., Craik, F. I. M., McIntosh, A. R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: A PET study of episodic memory. *Journal of Cognitive Neuroscience, 9,* 254–265.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cohen, G., & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology, 4,* 187–197.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities, mechanisms, and performances* (pp. 409–422). Amsterdam: Elsevier.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125,* 159–180.

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik

& T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 474–479.

Delaney, R. C., Rosen, A. D., Mattson, R. H., & Novelly, R. A. (1980). Memory function in focal epilepsy: A comparison of nonsurgical, unilateral temporal lobe and frontal lobe samples. *Cortex, 16,* 103–117.

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37,* 1189–1199.

Elfkides, A., Yiultsi, E., Kangellidou, T., Kounti, F., Dina, F., & Tsolaki, M. (2002). Weschler Memory Scale, Rivermead Behavioral Memory Test, and Everyday Memory Questionnaire in healthy adults and Alzheimer's patients. *European Journal of Psychological Assessment, 18,* 63–77.

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19,* 617–625.

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

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 691–702.

Harwood, E., & Naylor, G. F. K. (1969). Recall and recognition in elderly and young subjects. *Australian Journal of Psychology, 21,* 251–257.

Hayes, S. M., Ryan, L., Schnyer, D. M., & Nadel, L. (2004). An fMRI study of episodic memory: Retrieval of object, spatial, and temporal information. *Behavioral Neuroscience, 118,* 885–896.

Hermann, B. P., Wyler, A. R., Richey, E. T., & Rea, J. M. (1987). Memory function and verbal learning ability in patients with complex partial seizures of temporal lobe origin. *Epilepsia, 28,* 547–554.

Herrmann, D. J., Schooler, C., Caplan, L. J., Lipman, P. D., Grafman, J., Schoenbach, et al. (2000). The latent structure of memory: A confirmatory factor-analytic study of memory distinctions. *Multivariate Behavioral Research, 36,* 29–51.

Hirst, W., Johnson, M. K., Kim, J. K., Phelps, E. A., Risse, G., & Volpe, B. T. (1986). Recognition and recall in amnesics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 445–451.

Hirst, W., Johnson, M. K., Phelps, E. A., & Volpe, B. T. (1988). More on recognition and recall in amnesics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 758–762.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117–144.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant? A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1,* 179–188.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1–55.

Ichikawa, S. (1983). Verbal memory span, visual memory span, and their correlations with cognitive tasks. *Japanese Psychological Research, 25,* 173–180.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and Aging, 15,* 157–175.

Jonker, C., Geerlings, M. I., & Schmand, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population-based studies. *International Journal of Psychiatry, 15,* 983–991.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press.

Klauer, K. C., & Zhao, Z. (2004). Double dissociation in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 135,* 355–381.

Kohler, S., Moscovitch, M., Winocur, G., Houle, S., & McIntosh, A. R. (1998). Networks of domain-specific and general regions involved in episodic memory for spatial location and object identity. *Neuropsychologia, 36,* 129–142.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

La Voie, D., & Light, L. L. (1994). Adult age differences in repetition priming: A meta-analysis. *Psychology and Aging, 9,* 539–553.

Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 531–541.

Loring, D. W., Lee, G. P., Martin, R. C., & Meador, K. J. (1988). Material-specific learning in patients with partial complex seizures of temporal lobe origin: Convergent validation of memory constructs. *Journal of Epilepsy, 1,* 53–59.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

Mecklinger, A., & Meinshausen, R.-M. (1998). Recognition memory for object form and object location: An event-related potential study. *Memory & Cognition, 26,* 1068–1088.

Milner, B. (1975). Psychological aspects of focal epilepsy and its neurosurgical management. *Advances in Neurology, 8,* 299–321.

Mitchell, D. B., & Bruss, P. J. (2003). Age differences in implicit memory: Conceptual, perceptual, or methodological? *Psychology and Aging, 18,* 807–822.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 269–310). Cambridge, MA: MIT Press.

Nyberg, L. (1994). A structural equation modeling approach to the multiple memory systems question. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 485–491.

Nyberg, L., Maitland, S. B., Ronnlund, M., Backman, L., Dixon, R., Wahlin, A., & Nilsson, L.-G. (2003). Selective adult age differences in age-invariant multifactor model of declarative memory. *Psychology and Aging, 18,* 149–160.

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1185–1191.

Parker, E. S., Landau, S. M., Whipple, S. C., & Schwartz, B. L. (2004). Aging, recall, and recognition: A study on the sensitivity of the University of Southern California Repeatable Episodic Memory Test (USC-REMT). *Journal of Clinical and Experimental Neuropsychology, 26,* 428–440.

Pezdek, K. (1983). Memory for items and their spatial locations by young and elderly adults. *Developmental Psychology, 19,* 895–900.

Postman, L. (1963). One trial learning. In C. F. Cofer & B. S. Musgrave

(Eds.), *Verbal learning and verbal behavior* (pp. 295–332). New York: McGraw-Hill.

Salthouse, T. A. (1974). Using selective interference to investigate spatial memory representations. *Memory & Cognition, 2,* 749–757.

Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study.* New York: Cambridge University Press.

Schonfield, D., & Robertson, B.-A. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20,* 228–236.

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

Shimamura, A. P. (1995). Memory and frontal lobe function. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 803–813). Cambridge, MA: MIT Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Squire, L. R., & Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus, 18,* 205–211.

Taub, G. E., McGrew, K. S., & Witta, E. L. (2004). A confirmatory analysis of the factor structure and cross-age invariance of the Wechsler Adult Intelligence Scale–III. *Psychological Assessment, 16,* 85–89.

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General, 107,* 393–419.

Warrington, E. K., & Sanders, H. I. (1970). The fate of old memories. *Quarterly Journal of Experimental Psychology, 23,* 432–442.

Wechsler, D. (1997a). *WAIS–III–WMS–III technical manual*. San Antonio, TX: The Psychological Corporation.

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

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

Whiting, W. L., & Smith, A. D. (1997). Differential age-related processing limitations in recall and recognition tasks. *Psychology and Aging, 12,* 216–224.

Widaman, K. F. (1992). Multitrait–multimethod models in aging research. *Experimental Aging Research, 18,* 185–201.

Wilde, N. J., Strauss, E., Chelune, G. J., Hermann, B. P., Hunter, M., Loring, D. W., et al. (2003). Confirmatory factor analysis of the WMS–III in patients with temporal lobe epilepsy. *Psychological Assessment, 15,* 56–63.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.

Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin & Review, 9,* 575–582.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (1999). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Erlbaum.

Zelinski, E. M., & Lewis, K. L. (2003). Adult age differences in multiple cognitive functions: Differentiation, dedifferentiation, or process-specific change? *Psychology and Aging, 18,* 727–745.

Received June 12, 2006
Revision received October 20, 2006
Accepted January 18, 2007 ■