Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 55 (2006) 102–125

Journal of Memory and Language

www.elsevier.com/locate/jml

An individual differences analysis of memory control

Timothy A. Salthouse *, Karen L. Siedlecki, Lacy E. Krueger

Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

Received 31 October 2005; revision received 15 March 2006
Available online 12 May 2006

Abstract

Performance on a wide variety of memory tasks can be hypothesized to be influenced by processes associated with controlling the contents of memory. In this project 328 adults ranging from 18 to 93 years of age performed six tasks (e.g., multiple trial recall with an interpolated interference list, directed forgetting, proactive interference, and retrieval inhibition) postulated to yield measures of the effectiveness of memory control. Although most of the patterns from earlier studies were replicated, only a few of the measures of memory control were reliable at the level of individual differences. Furthermore, the memory control measures had very weak relations with the age of the participant. Analyses examining the relations between established cognitive abilities and variables from the experimental tasks revealed that most of the variables were related only to episodic memory ability.

Keywords: Control of memory contents; Aging; Individual differences; Memory processing

There are several reasons to expect the ability to efficiently control the contents of one's memory to be related to performance in a variety of memory tasks. For example, it is sometimes desirable to discard information from memory when it is no longer relevant or has been found to be invalid, or to maintain previously presented information in an accessible state during the processing of other information. Individuals with greater control over the contents of their memory might therefore be expected to be better at discriminating between to-be-remembered and to-be-forgotten items, less susceptible to proactive interference, less susceptible to inhibition associated with the retrieval of related items, and in general have a better overall memory. The current project was designed to investigate these expectations in the context of two major questions.

First, to what extent are measures of memory control related to one another and to other cognitive abilities? The rationale is that if there is a distinct dimension of human functioning associated with controlling the contents of one's memory then different measures of the efficiency of controlling memory would be expected to be at least moderately correlated with one another, but only weakly correlated with measures of other cognitive abilities. One way of representing this hypothesis is portrayed in Fig. 1. Notice that variables assumed to reflect different types of memory control are postulated to be related to one another such that they can be inferred to represent the same theoretical construct, and that construct is expected to have only weak relations with other constructs, indicating that it represents something different from what is represented by those constructs. Both of these conditions need to be satisfied to establish the existence of construct validity because moderate to strong relations among the variables are necessary to exhibit convergent validity, but the

* Corresponding author.
E-mail address: salthouse@virginia.edu (T.A. Salthouse).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2006.03.006

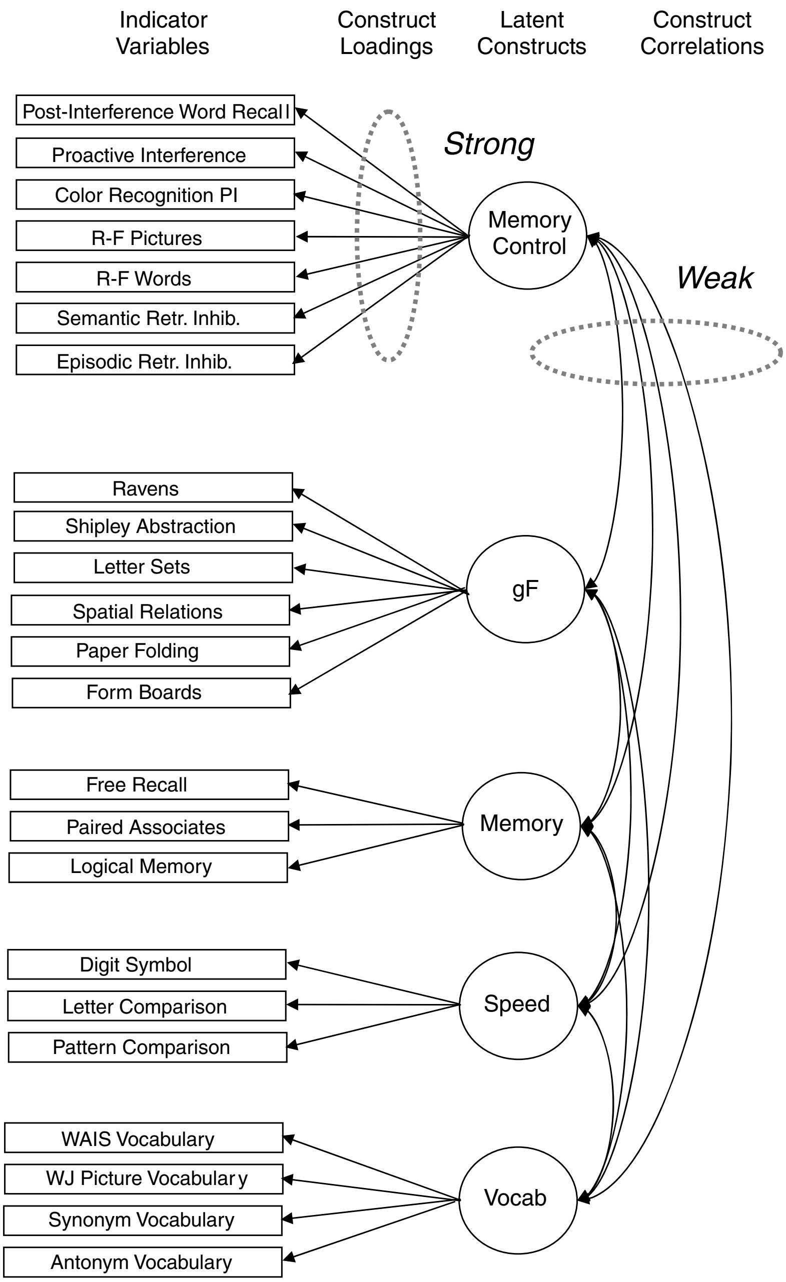

Fig. 1. Model illustrating the analytical approach used to investigate the construct validity of a construct concerned with controlling the contents of memory.

relations of that construct to other constructs should not be very high to ensure that the construct is distinct, and exhibits discriminant validity. Rigorous examination of construct validity is desirable because there has been a proliferation of references to different theoretical constructs with little or no investigation of the relation of those constructs to existing constructs. One of the limitations of this practice is that individuals may be characterized as varying in a single theoretically interesting dimension, such as executive functioning or working memory, when they may also vary in other dimensions that could be contributing to any relations that might be observed. A desirable first step in establishing that a theoretical construct corresponds to something distinct from already established constructs is evidence of this type of convergent and discriminant validity.

The second major question addressed in the project was what are the effects of aging on various measures of memory control? Age-related declines in memory are well documented (e.g., Zacks, Hasher, & Li, 2000), and thus increased age might be hypothesized to be associated with poorer control over the contents of memory. However, because a great many cognitive variables have been found to be related to age (e.g., Craik & Salthouse, 2000; Salthouse, 2004a, 2004b), age-related effects on measures of memory control might not be independent of age-related influences on other variables. To the extent that this is the case, no separate explanation of age-related effects on the memory control measures may be needed. In order to investigate this possibility, age-related effects on the target variables should be examined in the context of age-related effects on variables that have already been established to be related to age.

Fig. 2 illustrates the analytical approach employed in the current study to conduct this type of contextual analysis. The organization of other variables into cognitive abilities is portrayed on the left and the solid arrows from age to the abilities represent relations that have been well established in past research (e.g., Salthouse, 2004b, 2005). The key information in this type of analytical model concerns the strengths of the dotted lines to the target (i.e., memory control) variable. The relation from age represents the direct influence associated with increased age that is independent of any age-related effects on the variable operating through the cognitive abilities. Although a very large number of variables might be found to be related to age, only if this unique relation was significantly different from zero would a separate explanation necessarily be required for the age-related influences on the target variable.

Relations from the cognitive abilities in this analysis reflect the influence of each of those abilities on the target variable, and thus can be informative about what the variable represents. That is, a variable can be inferred to involve primarily memory processes if it has a strong relation with memory ability but weak relations with other abilities. In contrast, the variable can be inferred to largely represent fluid intelligence (gF) if the only strong relation it has is with gF ability. Note that in order to be informative about the relative contributions of different constructs the analysis must include multiple constructs. When only a single hypothesized mediator construct is included in the analysis the estimates can be distorted because all of the influences that are shared among different constructs are absorbed by the one construct included in the analysis.

Interpretation of results from this type of contextual analysis depends on the constructs used as the context, and the particular set of contextual constructs considered most meaningful depends on one's theoretical perspective. Nevertheless, the analytical procedure portrayed in Fig. 2 provides an objective method of investigating the meaning of a variable in terms of the relative strengths of different types of influences. Moreover, Salthouse (2005) recently provided evidence for the validity of this procedure by demonstrating that, as one would expect, variables selected to represent each reference construct had strong relations only from that ability construct and little or no relations with other ability constructs. The critical question with respect to the memory control measures is whether they represent something more than episodic memory, in which case they would be predicted to have influences from other constructs.

Selection of the tasks and variables is critical in research attempting to investigate construct validity because there must be a plausible basis to expect that the measures all reflect the relevant theoretical concept. It was fairly easy to identify candidate measures in recent projects investigating constructs of executive functioning, inhibition, time-sharing, and updating (Salthouse, Atkinson, & Berish, 2003), prospective memory (Salthouse, Berish, & Siedlecki, 2004), and source memory (Siedlecki, Salthouse, & Berish, 2005). Unlike the situation with these other constructs, however, we were unable to locate explicit references to a single construct concerned with controlling the contents of one's memory. Nevertheless, a number of researchers have speculated about the role of attention, executive functioning, or other types of control processes in resisting interference, suppressing information that is no longer relevant, and inhibiting possible competitors at the time of retrieval (e.g., Anderson, 2003; Bjork, 1998; Engle & Kane, 2004; Kane & Engle, 2000). To the extent that various measures hypothesized to reflect memory control have a common influence, such as executive processes or controlled attention, they can be expected to represent the same dimension of human functioning. Research from an individual differences perspective can therefore be valuable in determining whether there is commonality among different methods of operationalizing a theoretical construct concerned with memory control. It is important to note that the proposed analytical methods do not force the different variables to be grouped together, but rather provide a means of examining the empirical strength of any relations that might exist among the variables. These techniques therefore allow one to investigate whether it is meaningful, at least from an individual differences perspective, to use the same label, such as memory control, to refer to variables from different types of tasks. It is desirable that research using these types of analytical procedures include a broad variety of candidate variables because theorists are likely to differ with respect to the particular variables that would be predicted to provide the best reflections of a construct such as memory control. However, an advantage of analytical methods such as these is that one can rely on objective empirical information to deter-

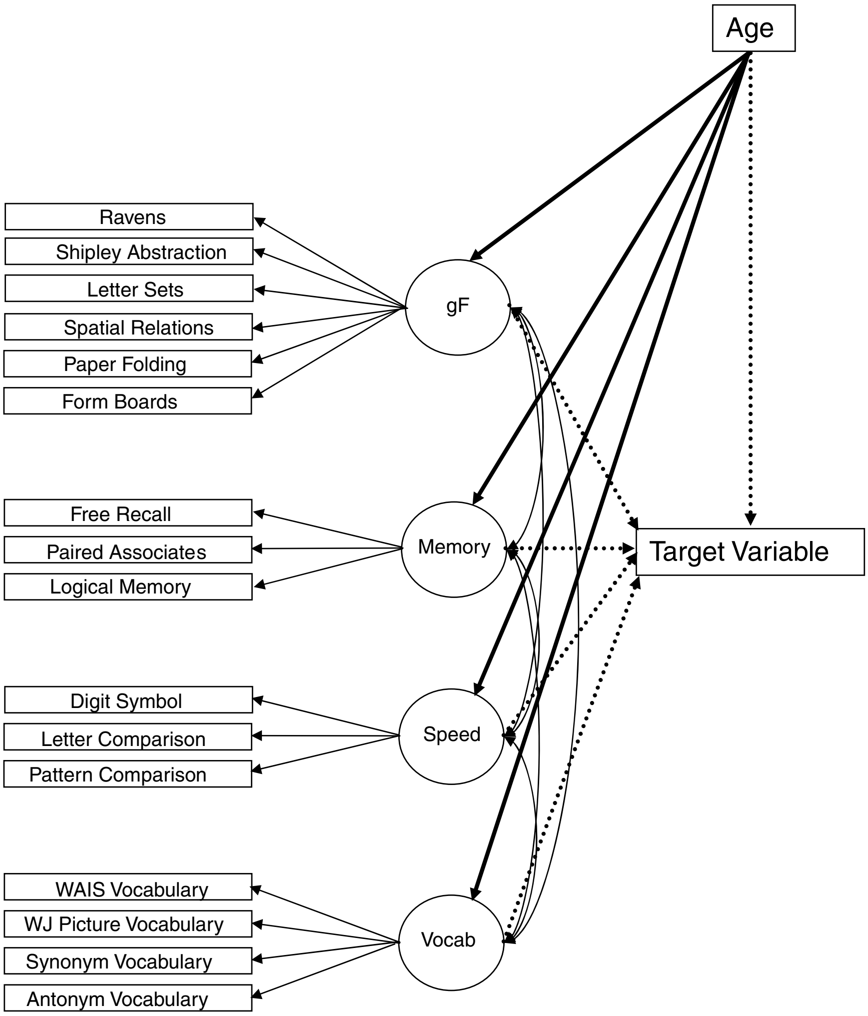

Fig. 2. Model illustrating the analytical approach used to investigate the relations of different cognitive abilities and of age on the target variables.

mine whether different combinations of variables reflect the same dimension of individual differences.

Six tasks were used in the current study to assess effectiveness of controlling the contents of one's memory. In the Multiple Trial Word Recall task the same list of 12 unrelated words is presented four times with a recall attempt after each presentation, followed by the presentation and attempted recall of a different list of words, and then an attempt to recall words from the original list without seeing or hearing them again. Better resistance to interference, and hence control of the contents of one's memory, can be postulated to be manifested in a smaller difference in recall before and after the attempt to recall the words from the other list.

Dunlosky and Salthouse (1996) reported analyses of the acquisition and forgetting of individual items in multiple trial free recall and found nearly parallel functions across trials for adults of different ages. However, the difference between recall in the pre- and post-interference trials was significantly larger with increased age, which implies that susceptibility to this type of interference may increase, and effectiveness of the relevant form of memory control may decrease, with increased age.

The Directed Forgetting task consisted of the presentation of an intermixed set of words and pictures with an instruction immediately after each item indicating that it should either be remembered or forgotten. At the end of the list the participants are asked to recall all items, including those they were initially instructed to forget. Because the remember or forget instruction does not occur until after the presentation of the item, in order for an item to be remembered later the participant must engage in at least some encoding while the stimulus is presented. The initial encoding could be rather superficial, perhaps only serving to maintain the item (Woodward & Bjork, 1971), but it is presumably the same for both to-be-remembered and to-be-forgotten items because the instruction about the fate of the item does not occur until after the presentation of the item. The directed forgetting procedure is therefore relevant to memory control because individuals who are effective at purging recently attended information from memory, or selectively rehearsing critical information, would be expected to have a large discrepancy between recall of to-be-remembered items and recall of to-be-forgotten items.

Results from previous studies investigating adult age differences in directed forgetting tasks have been mixed as there have been some reports of smaller discrepancies between the recall of to-be-remembered items and to-be-forgotten items with increased age, and some reports of no differences. However, a consistent pattern in most of the studies was that the age differences were larger for the to-be-remembered items than for the to-be-forgotten items (Dulaney, Marks, & Link, 2004; Earles & Kersten, 2002; Gamboz & Russo, 2002; Zacks, Radvansky, & Hasher, 1996), which could imply that increased age impairs the ability to remember information but may have little impact on the effectiveness of inhibiting to-be-forgotten information.

Progressively lower memory performance on successive lists of words involving the same type of information is known as proactive interference (PI) because it is assumed to reflect the interfering effects associated with prior exposure to similar material. One task used to assess PI in the current study involved the presentation of words from the same semantic category on three successive lists, with the participant asked to engage in a distracting activity prior to recall of each list, followed by a shift to words from a different semantic category for the next three lists. It has been suggested that individuals with better control of memory may be more effective at discarding old information when it is no longer relevant, and thus would be less susceptible to PI (e.g., Engle & Kane, 2004; Kane & Engle, 2000).

Quite a few studies have compared adults of different ages in measures of proactive interference (e.g., see reviews in Kausler, 1994 & Salthouse, 1991). Unfortunately, the results have not been very consistent as nearly as many studies have reported no differences as have reported differences, and the reasons for the discrepancies are not obvious. Because the sample sizes in most of the studies were relatively small, some of the inconsistency may have been attributable to chance fluctuation and low power and precision.

Siedlecki et al. (2005) recently reported progressively lower performance across successive lists in a task in which participants were asked to specify the color in which to-be-remembered words had been presented. The decrease in accuracy of color recognition across the successive lists is presumably attributable to proactive interference because the same colors were used with different words, and thus an individual who is better able to control the contents of his or her memory might be expected to exhibit less interference across successive lists.

Because the Siedlecki et al. (2005) study was conducted for a different purpose, the data were re-analyzed to examine possible age differences in this measure of proactive interference. Color recognition accuracies in the first and fourth lists were, respectively, 83 and .61 for adults age 18–39, .80, and .58 for adults age 40–59, and .72 and .52 for adults age 60–89. The lower accuracy in the fourth list compared to the first list is evidence for PI, but the similar reduction for each age group, and the correlation of −.00 between age and the difference in performance between list 1 and list 4, implies that there was no relation between age and this measure of PI.

Considerable research in the last 10–15 years has investigated a phenomenon known as retrieval-induced forgetting (e.g., Anderson, 2003; Bjork, 1998). A popular paradigm for investigating this retrieval inhibition phenomenon involves having participants initially study exemplars from several categories, followed by practice attempting to recall some of the exemplars from some of the categories. Either immediately or after a short delay the participants are then tested for their memory of all of the initially studied items. Members of the practiced categories that were not practiced are often not remembered as well as members of not-practiced categories (e.g., see Anderson, 2003, for a recent review), and this difference in memorability has been interpreted as reflecting a process of inhibition of the retrieval of non-practiced members of the practiced categories.

Two versions of the retrieval inhibition task were used in the current project. One was semantic, involving pre-existing category-exemplar pairs as the stimulus material, and the other was episodic, in which participants learned new associations between shapes (analogous to categories) and colors (analogous to exemplars). It is not surprising that the items that receive additional retrieval practice (i.e., RP+) often have superior memory performance relative to items from categories that are not practiced (i.e., NP). However, the comparison of primary interest was between recall of unpracticed exemplars from the practiced categories (RP−) and recall of exemplars from not-practiced categories (NP). Better control of the contents of one's memory, in the form of less inhibition associated with the retrieval practice, can be hypothesized to be manifested as a smaller NP-minus-RP− difference. We are not aware of any published reports investigating adult age differences in the effectiveness of this type of retrieval inhibition.

It is reasonable to ask whether the procedures described above all involve the same type of control over the contents of memory. At first impression the measures might appear to have little in common because some of the effects could be assumed to originate during presentation or encoding and others during test or retrieval, and some of the tasks appear to involve an active and deliberate type of control whereas any control involved in the other tasks might be argued to be more passive and automatic. For instance, directed forgetting with item cueing likely occurs immediately after encoding by active rehearsal of to-be-remembered items, perhaps accompanied by deliberate suppression of to-be-forgotten items. In contrast, retrieval inhibition is hypothesized to occur when potential competitors are inhibited during retrieval, and this suppression may occur relatively automatically and without conscious effort. Nonetheless, each of the phenomena described above can be hypothesized to be at least partially dependent on the effectiveness of either preserving or discarding information in memory. That is, item-cued directed forgetting may be based on differential rehearsal resulting in the strengthening of to-be-remembered items or the weakening of to-be-forgotten items, and in the interpolated word list and proactive interference paradigms less interference would be expected if the relevant information in memory could be preserved during other processing. Finally, control may be involved in the retrieval inhibition tasks in the form of resisting suppression of potential competitors at the time of retrieval. Theoretical arguments could almost certainly be developed to predict that some of the memory control measures might be more closely related to one another than others. However, by including a broad variety of measures in the same analysis we are able to rely on empirical methods to simultaneously investigate the plausibility of any number of hypothesized clusters of measures. That is, correlations provide objective information about the degree of relatedness among variables and thus they can be used to determine which groupings of variables are most meaningful.

The analyses investigating the models portrayed in Figs. 1 and 2 are based on correlations, and thus it is important to consider two prerequisites for the meaningful interpretation of correlations; moderately large sample size, and adequate reliability of the critical measures. Size of the sample is relevant because it is directly related to the precision of the estimates. That is, when the sample size is small the confidence intervals around the estimates are large, and consequently precision is low. To illustrate, with a sample of 300 the 95% confidence interval around a correlation of .50 is relatively narrow (i.e., 41 to .58), but the interval spans a much broader range (i.e., .17 to .73) when the sample is only 30.

Reliability must be considered when interpreting correlations because it represents the proportion of systematic (non-error) variance in the measure that could be associated with other variables, and thus it sets an upper limit on potential correlations of the measure with any other variable. The possibility that correlations are low because of weak reliability of the measures must therefore be ruled out before attempting to interpret an apparent lack of a relation.

The issue of reliability is particularly important when the measures of interest are derived from a contrast between levels of performance in two or more conditions, which is often the case in cognitive research attempting to isolate specific theoretical processes. The concern is that the effect could be highly significant when everyone exhibits the effect to nearly the same degree, but if there is little variability across people in the magnitude of the difference then the measure of the effect will not be reliable. A paradoxical situation can therefore arise in which an effect is robust and statistically significant in experimental research because the between-person variability in the relevant measure is relatively small, but the small between-person variability may be associated with low reliability in individual difference research, which will limit possible correlations of the measure with other variables.

To summarize, the primary purpose of the present project was to investigate relations of age and cognitive abilities on measures hypothesized to reflect the efficiency of controlling the contents of memory. One analytical approach relied on the model represented in Fig. 1 to determine whether it is meaningful to treat different measures of memory control as reflections of a distinct cognitive ability. To the extent that people vary in the ability to exert control over their memory in different types of tasks, the memory control variables should be significantly correlated with one another, but if memory control is a distinct ability it should have only moderate correlations with other cognitive abilities. A second analytical approach used the model in Fig. 2 to investigate the relative contribution of different cognitive abilities, and the uniqueness of age-related influences, on measures of memory control. The rationale is that the strengths of the relations from the ability constructs to the target variable are informative about the relative influences of these abilities on the variable, and the direct relation from age is informative about age-related influences on the target variable that are statistically independent of influences on the abilities. The study involved a moderately large sample of participants, and estimates of reliability were obtained for most of the target variables.

Method

Participants

The participants consisted of 328 adults between 18 and 93 years of age who reported to the laboratory for three sessions of approximately two hours each. Only a few college students participated in the project, and newspaper ads, flyers, and referrals from other participants were used to recruit participants, who received $120 as compensation. Characteristics of the participants, with the sample divided into three age groups for ease of description, are summarized in Table 1. It can be seen that the participants averaged close to 16 years of education, and that amount of education was not significantly related to age. Most participants reported themselves to be in very good to excellent health, and only one individual reported his or her health to be poor. All of the participants had scores above 23 on the Mini Mental Status Exam (Folstein, Folstein, & McHugh, 1975) often used as a preliminary screening instrument for dementia. Different types of tasks were performed on each session, and short breaks were provided between tasks whenever desired by the participant.

The representativeness of the sample, both with respect to the US population and relative to participants of different ages, can be evaluated by examining age-adjusted scaled scores from standardized tests in the Wechsler Adult Intelligence Scale III (Wechsler, 1997a) and the Wechsler Memory Scale III (Wechsler, 1997b) test batteries. These scaled scores have a mean of 10 and a standard deviation of 3 in the nationally representative samples used to establish the norms for the tests. The means in this sample ranged from 10.6 to 13.6, indicating that most of the participants were functioning well above the national averages. However, there were only weak relations between age and the scaled scores, which suggests that participants of different ages in the sample were equally select relative to their age groups. The only significant relation between age and a scaled score was positive, suggesting that with increased age participants in the sample may have been functioning at a somewhat higher level relative to their age peers from the normative sample in memory for meaningful material (i.e., Logical Memory).

Procedure

The tests were administered across three separate sessions that most participants completed within a 2-week interval. Each session lasted about two hours, and involved a variety of tasks that required between 5 and 20 min each. The order of test presentation, which was the same for all participants, consisted of WAIS Vocabulary, WJ Picture Vocabulary, Digit Symbol, Multiple Trial Word Recall, Pattern Comparison, Letter Comparison, Shipley Abstraction, and Paired Associates on Session 1, Brown–Peterson Interference, Antonym–Synonym Vocabulary, Ravens, Paper Folding, Episodic Retrieval Inhibition, and Form Boards on Session 2, and Letter Sets, Semantic Retrieval Inhibition, Spatial Relations, Logical Memory, Color Word Proactive Interference, and Directed Forgetting for Words and Pictures on Session 3. Several of the tests (e.g., WAIS Vocabulary, WJ Picture Vocabulary) were discontinued if the participant was incorrect on a specified number of consecutive trials, and others had time limits to ensure efficient assessment (e.g., 120 s for Digit Symbol, 30 s for each of two trials in the Pattern Comparison and Letter Comparison tests, 5 min for Shipley Abstraction and Antonym–Synonym Vocabulary, 8 min for Form Boards, and 10 min for Ravens, Paper Folding, Spatial Relations, and Letter Sets).

Table 1
Sample characteristics

	18–39		40–59		60–93		Age corr.
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
N	89		133		106		—
Age	28.3	6.0	49.7	5.4	70.8	7.7	—
Prop. female	.66		.66		.60		−.04
Health	1.9	0.8	2.1	0.8	2.3	0.9	.21*
Yrs. education	15.2	2.2	15.8	2.8	15.9	2.6	.13
MMSE	29.2	1.2	28.8	1.4	28.5	2.0	−.18*
Scaled scores							
Vocabulary	12.8	3.1	12.2	3.0	13.6	2.9	.10
Digit symbol	11.0	2.8	10.6	2.8	11.4	2.7	.09
Logical memory	11.9	2.5	11.9	2.3	13.0	2.8	.16*
Word recall	11.8	3.2	11.7	3.4	12.4	3.3	.08

Note. Health was a self rating on a scale ranging from 1 for excellent to 5 for poor. MMSE is the Mini Mental Status Exam (Folstein et al., 1975), and scaled scores are age-adjusted scores relative to the nationally representative normative sample in the WAIS III (Wechsler, 1997a) and WMS III (Wechsler, 1997b).

* $p < .01$.

Cognitive reference tests

The tests used to assess established cognitive abilities are briefly described in Table 2 and summary statistics for the relevant variables are presented in Table 3. Inspection of Table 3 reveals that all variables had good reliability, and most were significantly related to age. Eight of the 5248 cells (i.e., 16 variables × 328 participants) had missing data and in order to simplify the analyses the missing values were replaced with the variable mean. (Nearly identical results were obtained when a maximum likelihood estimation algorithm was used to deal with missing data rather than mean substitution.)

A confirmatory factor analysis on these reference variables indicated that, as found in other data sets (e.g., Salthouse, 2004b, 2005), the variables could be interpreted as representing five correlated cognitive abilities. However, the reasoning and spatial visualization factors were highly correlated with one another ($r = .84$), which led to problems of multicollinearity in some of the analyses. These two factors were therefore combined into a single fluid ability (gF) factor in all of the remaining analyses. Standardized coefficients indicating the relations of variables to the ability factors, and of the factors to one another, are also presented in Table 3. Notice that all of the variables had moderately high loadings on their respective factors and that the fit of the model was respectable despite the simple structure in which each variable was related to only one ability construct.

To portray the age relations on the abilities, composite scores were formed by averaging the z scores for the variables representing each ability construct, and plotting the mean composite scores as a function of age in Fig. 3. The figure exhibits the typical pattern of nearly continuous decreases with age in most cognitive abilities, with the exception of vocabulary which increases until about age 50 or 60 (e.g., Salthouse, 2004a, 2004b). Because they represent well-documented findings, the results in Table 3 and Fig. 3 can be considered to be the context within which age-related influences on new variables should be examined.

Memory control tasks

Multiple trial word recall

On the first session of the project the participants performed the Word List Recall test from the Wechsler

Table 2
Description of cognitive variables included in the analyses

Variable	Description	Source
Ravens	Determine which pattern best completes the missing cell in a matrix	Raven (1962)
Shipley abstraction	Determine the words or numbers that are the best continuation of a sequence	Zachary (1986)
Letter sets	Identify which of five groups of letters is different from the others	Ekstrom et al. (1976)
Spatial relations	Determine the correspondence between a 3D figure and alternative 2D figures	Bennett et al. (1997)
Paper folding	Determine the pattern of holes that would result from a sequence of folds and a punch through folded paper	Ekstrom et al. (1976)
Form boards	Determine which combinations of shapes are needed to fill a larger shape	Ekstrom et al. (1976)
Logical memory	Number of idea units recalled across three stories	Wechsler (1997b)
Free recall	Number of words recalled across trials 1–4 of a word list	Wechsler (1997b)
Paired associates	Number of response terms recalled when presented with a stimulus term	Salthouse et al. (1996)
Digit symbol	Use a code table to write the correct symbol below each digit	Wechsler (1997a)
Letter comparison	Same/different comparison of pairs of letter strings	Salthouse and Babcock (1991)
Pattern comparison	Same/different comparison of pairs of line patterns	Salthouse and Babcock (1991)
WAIS vocabulary	Provide definitions of words	Wechsler (1997a)
WJ picture vocabulary	Name the pictured object	Woodcock and Johnson (1990)
Antonym vocabulary	Select the best antonym of the target word	Salthouse (1993)
Synonym vocabulary	Select the best synonym of the target word	Salthouse (1993)

Table 3
Estimated reliabilities (Rel), means, standard deviations (*SD*s), age correlations (Age *r*), standardized factor loadings, and factor correlations for the 16 reference variables

	Rel	Mean	*SD*	Age *r*	Factors			
					gF	Mem	Spd	Voc
Variable								
Ravens	.78	7.4	3.5	−.47*	.85+			
Shipley abstraction	.84	12.9	3.6	−.37*	.82*			
Letter sets	.74	11.0	2.8	−.32*	.72*			
Spatial relations	.88	8.8	5.2	−.37*	.80*			
Paper folding	.77	6.4	2.9	−.30*	.81*			
Form boards	.85	7.2	4.4	−.45*	.77*			
Logical memory	.80	45.6	9.9	−.23*		.65+		
Free recall	.88	34.8	6.3	−.36*		.78*		
Paired associates	.76	2.9	1.7	−.35*		.77*		
Digit symbol	.93	70.9	17.5	−.57*			.70+	
Letter comparison	.87	10.5	2.5	−.46*			.66*	
Pattern comparison	.90	16.6	3.7	−.59*			.85*	
WJ picture vocab.	.86	18.6	5.0	.31*				.82+
Synonym vocab.	.83	6.7	3.0	.29*				.89*
Antonym vocab.	.82	6.1	3.0	.17*				.83*
WAIS vocab.	.90	51.5	10.3	.12				.88*
Factor correlations								
gF					—	.67*	.66*	.48*
Mem						—	.49*	.38*
Speed							—	.07
Age					−.51*	−.48*	−.72*	.30*
Fit statistics								
X^2/df	343/98							
CFI	.92							
RMSEA	.09							

Note. The factors were fluid (gF), episodic memory (Mem), perceptual speed (Spd), and vocabulary (Voc). The "+" indicates that the unstandardized coefficient was fixed to 1 to identify the factor. Reliability of the Digit Symbol variable was based on the correlation with a parallel form in another sample of 143 adults (Salthouse et al., in press). The fit statistics are the Comparative Fit Index (CFI) and the Root Mean Squared Error of Approximation (RMSEA), with values of the former above .90 and values of the latter below .10 often considered to indicate good fit.

* $p < .01$.

Memory Scale III (Wechsler, 1997b). This task involves the presentation of four trials of the same list of 12 unrelated words in the same order, with immediate recall requested after each list. A new list of 12 words is then presented followed by immediate recall, and then participants are asked to recall as many words as possible from the original list without another presentation of the list. Because the interpolated list can be considered as interference for the retention of the original list, the measure of memory control in this task was the difference in recall on trials before and after the new list of words.

Directed forgetting

This task occurred in the third session of the project. It involved the visual presentation for 5 s each of an intermixed set of 30 words and 30 pictures, with each item immediately followed either by the word "Remember" (in green letters) or the word "Forget" (in red letters) for 2 s. If the word "Remember" was displayed the participants were instructed to continue trying to remember the item, but if the word "Forget" was displayed they were told that they could forget the item.

A practice list of six items, three with Remember instructions and three with Forget instructions, preceded the experimental list, and was followed immediately by instructions to recall the Remember items. The order of the Remember and Forget instructions in the experimental list was determined randomly with the constraint that one-half of the words and one-half of the pictures were followed by each type of instruction. The same random order was used for all participants. The words were 6- and 7-letter words from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982), and the 30 pictures of familiar animals or objects were from Cansino, Maquet, Dolan, and Rugg (2002). At the end

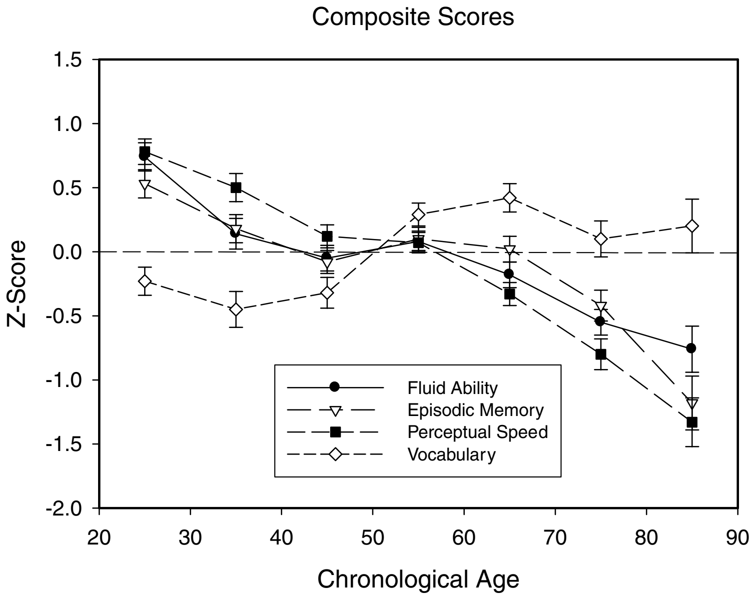

Fig. 3. Average *z*-scores for variables representing different cognitive abilities as a function of age. Bars around the data points are standard errors.

of the study list the examiner announced that all items should be recalled, even those initially presented with the "forget" instructions, and then recorded the items as they were orally recalled by the participant. The primary measures of memory control were the differences in recall between to-be-remembered and to-be-forgotten words and pictures.

Proactive interference

The proactive interference task was administered in the second session of the project. Six eight-word lists were visually presented on the computer display at a rate of 2 s per word, with each list followed by a three-digit number from which the participant was to count backwards by 3's for 16 s. An interval of 30 s was then allowed for the participant to orally recall as many words as possible from the immediately preceding list. The first three lists consisted of names of animals, and the next three lists consisted of names of countries. The animal names were obtained from Set 1 of Appendix A in Kane and Engle (2000), and most of the country names were obtained from the words in their Sets 3 and 4. The examiner recorded each word recalled by the participant on a response form, and also recorded the number of successful subtractions carried out during the 16 s interval. The measure of memory control was the difference between recall on the first and third lists involving words from the same semantic categories.

Color Word

The participants performed this task, which was identical to one included in Siedlecki et al. (2005), in the third session of the project. Words were presented on a computer screen, at a rate of one every 5 s, in one of four colors (i.e., blue, green, yellow, and red). The words were presented in four blocks of ten each, and all were five letters in length. Participants were instructed to pay attention to both the word and the color in which it appeared because they would be tested on both attributes. The participant attempted to write down all of the words from the previous list and then circled the name of the color in which the word had appeared. The measure of memory control was the difference in color recognition accuracy between the first and fourth lists.

Retrieval inhibition—Semantic

This task was based on a procedure described by Anderson, Bjork, and Bjork (1994) and was presented in the third session of the project. It involved four phases; study, retrieval practice, a 20-min interval occupied by performance of the Letter Sets and Spatial Relations tests, and a recall test.

In the study phase the participants were asked to study pairs of words, in which one word was a category label and the other word was a member of that category, so that they would be able to remember them later. Each of 60 category-exemplar pairs (i.e., 6 pairs from 10 categories) was displayed for 5 s. In order to minimize primacy and recency effects, the first two and the last two items in the study list were from categories (i.e., leather and hobbies) that were not included in either the retrieval practice phase or the test phase.

In the retrieval practice phase of the task displays of the category label and the first two letters of one of the category members presented earlier were shown for 10 s each, with the participant instructed to recall the name of the category member that had been presented in the previous study phase. Twelve of the original 60 items (i.e., three members each from the insects, fish, fruits, and weapons categories) were presented three times each, with the examiner recording the participant's responses. As in the studies by Anderson et al. (1994), no feedback was presented in the retrieval practice phase.

Stimuli in the test phase consisted of a category label and the first two letters of one of the category members. Each of 48 items (i.e., 12 RP+, 12 RP−, and 24 NP) was displayed for 10 sec with the examiner recording the participant's responses. The order of item type and category was balanced across the list, but the same order was used for all participants. The categories and the words listed as strong members of the categories in Appendix C of Anderson et al. (1994) were used as the stimulus materials. The measure of memory control was the difference between recall of NP items and RP− items.

At the end of the task the participants were asked if they attempted to integrate the words when they were trying to remember them. The question was identical to that used by Anderson and McCulloch (1999) and was as follows:

"During the study phase, when you saw a word pair, how often, if ever, did you intentionally think back to previously seen category members? For example, if you first saw Cars: Honda and then Cars: Ford, how often did you intentionally rehearse Ford and Honda together?"

The response was in the form of a frequency rating on a scale ranging from 1 for "never" to 4 for "almost always."

Retrieval inhibition—Episodic

This task was administered in the second session of the project. It was modified from a task described in Experiment 1 of Ciranni and Shimamura (1999),[1] which was designed as an episodic memory version of the Retrieval Inhibition paradigm because it is based on newly learned associations rather than pre-existing category-word associations. In our adaptation of the task the stimulus displays consisted of nine squares arranged in a circle. Each square contained one of three different shapes in one of nine different colors. Trials consisted of a probe item (i.e., a colored shape) in the middle of the circle, with the participant instructed to indicate the location of that colored shape by using the mouse to click on one of the nine squares on the perimeter of the circle. Feedback in the form of a display of the target (i.e., colored shape) in the correct location was presented immediately after each response. The trials were repeated until the participant achieved a criterion of three consecutive correct trials for each item, with items successively removed from the set as they met that criterion.

The participants next received three trials of practice retrieving the color for the shape in a particular location for four targets (i.e., two of the three colors for two of the three shapes). Probe stimuli consisted of an uncolored shape in one of the locations, with the participant attempting to select the appropriate color from a palette containing all nine colors. The correct color was displayed immediately after the response. The recognition test, which followed immediately, was similar to the retrieval practice phase except that no feedback was presented. The test stimuli consisted of the four previously practiced items, designated RP+, the two unpracticed colors from the practiced shapes, designated RP−, and three unpracticed shapes, classified NP. Two blocks, with different sets of shapes and colors used as stimuli in each block, were presented in immediate succession. The measure of memory control was the difference between recall of NP items and RP− items.

Results[2]

The results section is organized into three parts. The first part reports the results from the individual tasks used to assess efficiency of memory control. The second part summarizes results relevant to the investigation of the construct validity of a memory control construct based on the model in Fig. 1, and the third part describes the results of analyses investigating the application of the model in Fig. 2 to all of the measures from the experimental tasks.

Measures of memory control

Multiple trial word recall

Fig. 4 portrays the number of words recalled across successive lists in three age groups. The nearly parallel

[1] A sample of young adults in pilot research with the same task parameters as in Ciranni and Shimamura (1999) found the task to be very difficult and frustrating. We therefore tried to make the task easier by reducing the number of stimuli from 12 to 9 and decreasing the criterion for learning from seven correct trials per stimulus to three.

[2] Because of the moderately large sample size, a significance level of .01 was used in all statistical tests. Correlations with an absolute value greater than .14 are significantly different from zero with a sample of 328.

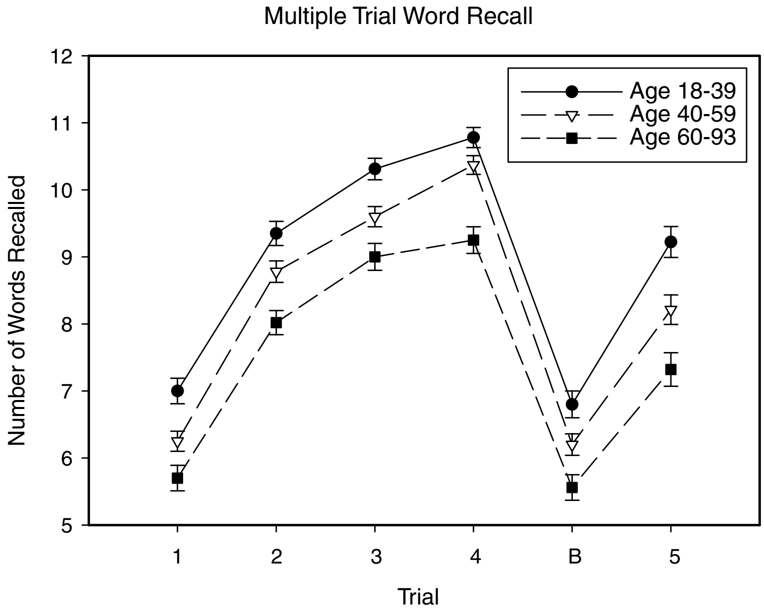

Fig. 4. Means (and standard errors) of recall on each trial involving the same (trials 1 through 4 and 5) or different (trial B) words in three age groups.

functions suggest that there was a main effect of age, but no differential effects of age either on the rate of learning across trials or on susceptibility to interference attributable to studying and recalling the interpolated list of different words.

The measure of memory control in this task was the difference in recall between trial 4 (pre-interference) and trial 5 (post-interference). The average difference of 2.0 was significantly greater than zero ($t = 20.4$), indicating that this pattern was characteristic of the overall sample. However, there was also variability in this difference (i.e., the standard deviation was 1.8), and thus the participants varied in the magnitude of the memory control measure.

No direct estimate of the reliability of the difference between recall on lists 4 and 5 was available, but it is noteworthy that the correlation between recall on the pre-interference (4) and the post-interference (5) trial is the same magnitude (i.e., 72) as the correlations between the preceding pairs of pre-interference trials (i.e., 1–2 = .67, 2–3 = .72, and 3–4 = .72). Because the rank-ordering of people is just as consistent before and after the interference word list as on successive trials before the interference list, the presentation of a different list of items does not appear to have any special significance in terms of individual differences.

Neither the correlation between age and the difference between trials 4 and 5 (i.e., $r = .12$), nor the semi-partial correlation between age and recall on trial 5 after controlling the variation in recall on trial 4 (i.e., −.08), was significantly different from zero. These results therefore suggest that there are little or no age-related effects on this measure of susceptibility to interference. An analysis of variance was conducted with trial (4 vs. 5) as a within-subjects factor and age group (i.e., ages 18–39, 40–59, and 60–93) as a between-subjects factor. The interaction in this analysis had an F-value of 3.65 ($p = .03$), suggesting that there was a trend for somewhat greater interference with increased age.

Directed forgetting

Fig. 5 portrays the means for the numbers of words and pictures recalled in the remember and forget conditions for the three age groups. There was a systematic age difference in the recall of to-be-remembered words ($r = -.39$) and in the recall of to-be-remembered pictures ($r = -.29$), and more to-be-remembered pictures were recalled than to-be-remembered words ($t = 11.5$). The age differences were not significant in the recall of to-be-forgotten words ($r = -.10$) or to-be-forgotten pictures ($r = .03$), but recall of the pictures that were not supposed to be remembered was greater than recall of words that were not supposed to be remembered ($t = 17.7$).

Reliabilities were estimated by treating scores on the first and second halves of the study trials as "items," and computing coefficient alpha. More to-be-remembered words were recalled from the first half of the list than from the second half, 2.5 vs. 1.8 ($t = 8.3$), but there was no difference from the first to the second half in

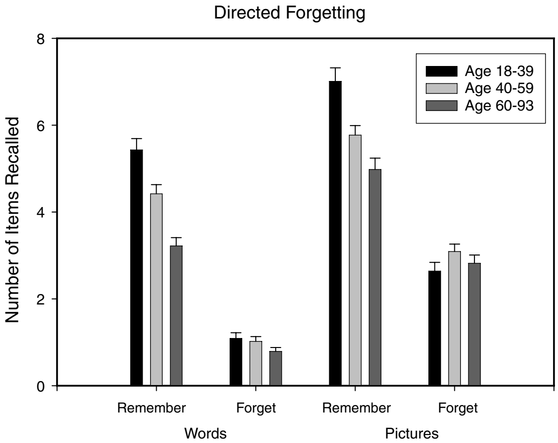

Fig. 5. Means (and standard errors) of recall for words and pictures that were to be remembered or were to be forgotten in three age groups.

the number of to-be-remembered pictures that were recalled, 3.0 vs. 2.9. More to-be-forgotten items were recalled from the second half of the study list for both words (i.e., .2 vs. .7, $t = 9.4$) and pictures (i.e., 1.0 vs. 1.9, $t = 9.9$).

The estimated reliabilities were .53 for to-be-remembered words (Word R), .17 for to-be-forgotten words (Word F), .49 for to-be-remembered pictures (Picture R), .38 for to-be-forgotten pictures (Picture F), .46 for the word R–F difference, and .41 for the picture R–F difference. Correlations between different measures, with the estimates adjusted for reliability in parentheses,[3] were Picture R–Word R = .57 (1.12), Picture F–Word F = .28 (1.10), Picture R–Picture F = .09 (.21), and Word R–Word F = .17 (.57). The weak correlations between the R and F measures indicate that it is not the case that people generally remember a greater or fewer number of items regardless of the remember or forget instructions. However, it is noteworthy that after adjusting for reliability there was nearly perfect overlap of the ordering of individuals in the picture R and word R measures, and in the picture F and word F measures, which suggests that the modality of presentation does not appear to be salient in terms of individual differences in this type of memory.

The primary measure of memory control in the directed forgetting task was the R-minus-F difference. For words this difference had a mean of 3.3, a standard deviation of 2.5, and a t value of 24.1, and the corresponding values for pictures were 3.0, 3.2, and 16.6. Both of the R–F scores were significantly different from zero, indicating that the average participant exhibited the pattern of greater recall of items that were to be remembered than of items that were to be forgotten. The standard deviations for the differences were also moderate, indicating that there was variability across people in the magnitude of the difference. As noted above, however, the estimated reliabilities of the differences were only .46 for words and .41 for pictures.

The word R–F difference had a correlation of −.33 with age, and the picture R–F difference had an age correlation of −.27. However, semi-partial correlations between age and the F score after control of the R score were only −.03 for words and .06 for pictures. This finding, together with the discovery that age differences were significant only for the R items and not the F items, implies that the age-related effects on the directed forgetting difference scores were largely attributable to effects on the to-be-remembered items.

Proactive interference

The number of words recalled in each list for the three age groups in the proactive interference task is plotted in the top panel of Fig. 6. The results reveal the typical decrease in recall from the first to the second list, and from the second to the third list, corresponding to the build-up of proactive interference. The level of recall increased when the items shifted to a different semantic category, reflecting a release from proactive interference, and then recall decreased from the fourth to the sixth lists as proactive interference increased again.

The functions in Fig. 6 are nearly parallel for different ages, which implies a similar build-up and release of proactive interference at each age. Because of the parallel functions, the same absolute magnitude difference corresponds to a larger proportion when the initial level of performance is lower. It is therefore possible that a relation apparent on the proportion measure could actually reflect a relation on the baseline, and for this reason the absolute difference measure was used in the current analyses.[4]

The average number of successful subtractions carried out during the retention interval was 6.4 (standard deviation of 2.8), and this number had a weak and non-significant relation to age ($r = -.13$). The correlation between the average number of words recalled and the average number of subtractions carried out during the retention intervals was .41. The positive correlation indicates that individuals with higher recall were

[3] The finding that the disattenuated correlations are greater than 1.0 may be attributable to underestimation of the actual reliability resulting in an inflation of the adjusted correlation.

[4] However, it should be noted that the correlation between the absolute (i.e., list 1 + 4 minus list 3 + 6) and proportional (i.e., [list 1 + 4 minus list 3 + 6]/list 1 + 4) measures of interference in this study was .90.

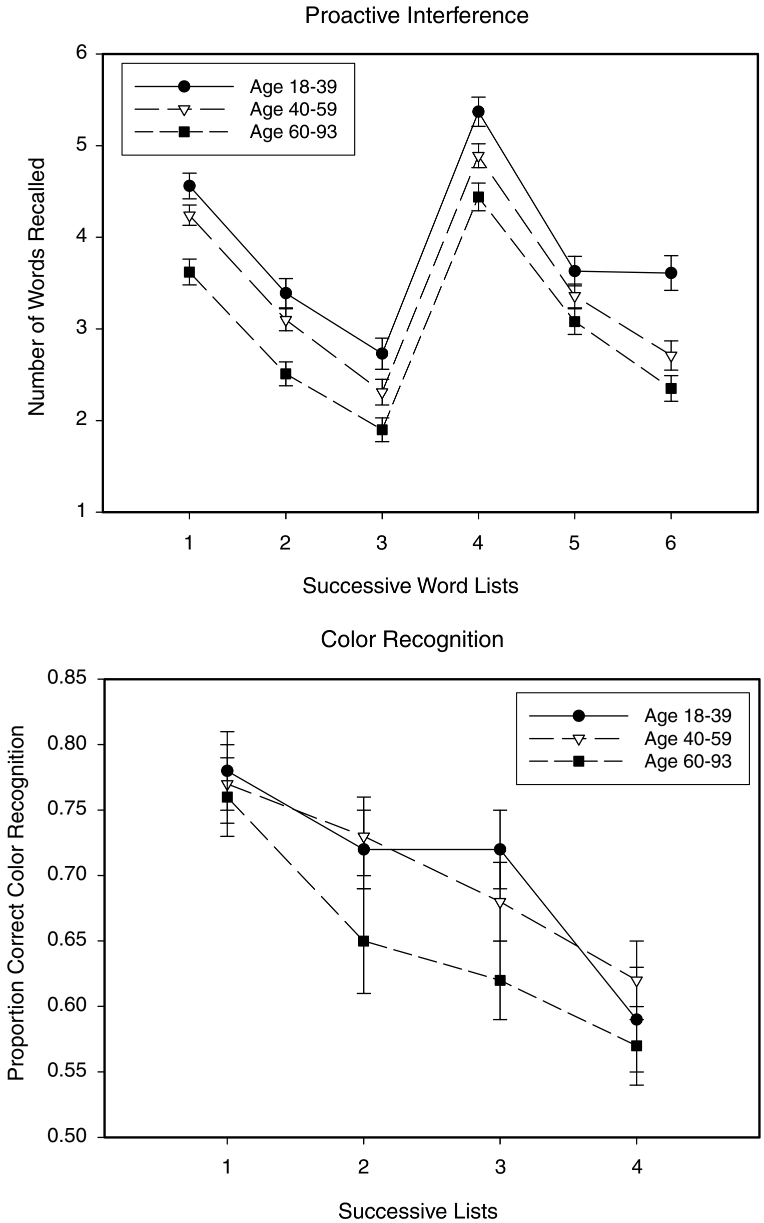

Fig. 6. Means (and standard errors) for three age groups across lists in the Brown–Peterson (top panel) and Color Word (bottom panel) tasks.

also faster at counting backwards, and therefore if anything, they were likely exposed to more distracting material than individuals who were slower at counting backwards.

The proactive interference measure of memory control was the average recall across lists 1 and 4 minus the average recall across lists 3 and 6. The mean difference was significantly different from zero (i.e., mean = 1.9, standard deviation = 1.3, $t = 27.4$), indicating that the average participant exhibited this pattern. However, the estimate of reliability for the difference between recall on trials 1 and 3 and between recall on trials 4 and 6, as computed with coefficient alpha based on the 1–3 and 4–6 differences, was only .20. The correlation of age with the 1 + 4 minus 3 + 6 difference was .01, and the semi-partial correlation between age and average recall on lists 3 and 6 after control of the variance in average recall on lists 1 and 4 was −.11.

Color word

Recall of words in the color word task was fairly constant across lists 1 through 4 as the averages were, respectively, 3.8, 3.8, 4.3, and 3.9. The age correlations were also similar, as they were −.36, −.38, −.32, and −.29 for lists 1 through 4, respectively.

The proportions of correct color attributions given that the word was recalled are portrayed in the bottom panel of Fig. 6 for three age groups. (Note that these are conditional measures because accuracy of the color assignment is only determined for words that were successfully recalled.) Although the functions are somewhat noisy, they appear to be nearly parallel and those for the three age groups are not very distinct.

The memory control measure was the color recognition score on list 1 minus the color recognition score on list 4. The mean difference was.18, the standard deviation was .43, and the t value was 7.4, indicating that most people exhibited this form of proactive interference. However, because participants differed in the number of words recalled on each list, it is difficult to compute an estimate of the reliability of the color word proactive interference measure. The correlation of age with the 1-minus-4 difference was .02, and the semi-partial correlation between age and score on list 4 after control of score on list 1 was −.07.

Semantic retrieval inhibition

The top panel of Fig. 7 displays the means in the three age groups of four measures of performance in the semantic retrieval inhibition task; the proportion correct for RP+ items in the retrieval practice phase, and the proportion correct for RP+, RP−, and NP items in the cued recall test. The highest average accuracy in the test phase was for RP+ items (.83), but there was no difference in accuracy between NP (.80) and RP− (.80) items. It is noteworthy that the level of accuracy for the RP+ items was about the same as in similar studies in Anderson et al. (1994) and Anderson and McCulloch (1999), but unlike those studies, in this study the accuracies for the NP and RP− items were similar to one another and only slightly lower than that for the RP+ items. Increased age was associated with significantly lower accuracy for RP+ items in the practice ($r = -.18$) and test ($r = -.17$) phases, but the correlations between age and the other measures were not significantly different from zero (i.e., −.09 for NP, and −.05 for RP−).

There was a significant negative correlation (i.e., $r = -.19$) between age and reported frequency of using an integration strategy, indicating that with increased age participants were less likely to report that they frequently attempted to integrate the study items. Mean

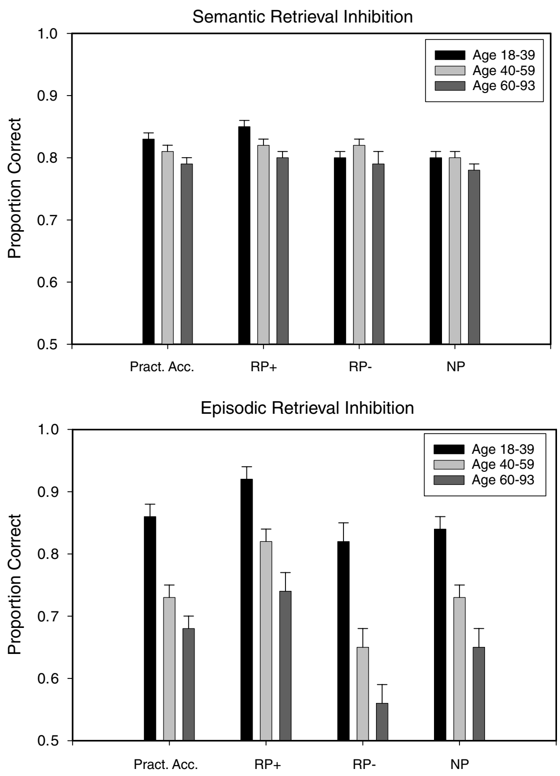

Fig. 7. Means (and standard errors) of cued recall accuracy for three age groups in the practice and test phases for different types of items in semantic (top panel) and episodic (bottom panel) retrieval inhibition tasks.

accuracy for each type of test item for individuals with different reported frequencies of integration of items during the study phase are reported in Table 4. (The total number of participants in this table was not equal to 328 because the question about integration strategy was mistakenly not asked of some participants.) Inspection of the table indicates that accuracy for each measure was somewhat higher among individuals reporting more frequent integration, but that all of the NP-minus-RP− difference scores were close to zero. These results are therefore inconsistent with the idea that amount of retrieval inhibition is related to the frequency of integrating the to-be-remembered items (e.g., Anderson, 2003; Anderson & McCulloch, 1999).

Accuracy was computed for each measure across the first 24 test trials and the second 24 trials in the cued recall test. Mean accuracy decreased from the first to the second half of the test for the RP+ items (i.e., 87 vs. .79, $t = 7.3$) and for the NP items (i.e., .88 vs. .72, $t = 19.5$), but it increased for the RP− items (i.e., .78 vs. .82, $t = 3.2$). The difference in accuracy between the NP and RP− trials was .09 on the first half of the trials and −.09 on the second half ($t = 13.5$), with both of

Table 4
Semantic retrieval inhibition results as a function of reported frequency of integration during study

Frequency	*N*	Pract.	RP+	RP−	NP	(NP − RP−)
Never	49	.77 (.14)	.78 (.14)	.80 (.22)	.74 (.15)	−.06 (.26)
Infrequent	71	.79 (.13)	.82 (.12)	.79 (.13)	.78 (.14)	−.01 (.16)
Frequent	107	.82 (.12)	.82 (.13)	.80 (.14)	.80 (.11)	−.00 (.14)
Almost always	78	.84 (.11)	.87 (.10)	.83 (.12)	.84 (.09)	.01 (.11)

Note. Numbers in parentheses are standard deviations.

these differences significantly different from zero (i.e., $t > 9.5$). Although the positive NP-minus-RP− difference for trials in the first half of the recognition test was consistent with the operation of retrieval inhibition, the overall pattern was rather unusual because accuracy on the NP trials was as high as on the RP+ trials, which indicates that there was no enhancement of performance associated with prior practice retrieving the items.

The memory control measure in the semantic retrieval inhibition task was the difference in accuracy between NP trials and RP− trials. The average difference was only −.01, with a standard deviation of .12, and it was not significantly differ from zero. The estimated reliability of the NP-minus-RP− difference based on the scores in the two halves of the test was actually negative (−.10). Age of the participant was not significantly correlated with this difference (i.e., −.02), or with the RP-measure after statistical control of the variation in the NP measure by means of semi-partial correlation (i.e., −.03). Retrieval practice resulted in a slight enhancement of recall because the RP+-minus-NP difference was significantly greater than zero (i.e., mean = .03, standard deviation = .15, $t = 3.4$), but this difference was not significantly related to age (i.e., $r = -.08$).

Episodic retrieval inhibition

The first phase of the episodic retrieval inhibition task involved learning associations between locations and colored shapes. Some participants found it very difficult to achieve the criterion of three successive trials correctly matching each of the nine colored shapes to its location because although the mean number of trials to criterion on the first set of trials was 54, the range was from 33 to 515. Many of the participants who experienced difficulty in this phase of the task were older adults as the correlation between age and the number of trials to criterion was .30.

The proportions of correct responses for the RP+ items in the practice phase and for the RP+, RP−, and NP items in the test phase for three age groups are displayed in the bottom of Fig. 7. Adults over the age of 60 had much lower accuracy for all measures than the other age groups, but none of the groups exhibited much difference across the measures.

Accuracies were also computed separately for the first and the second block of test trials. There was no significant difference in NP accuracy from the first to the second block (i.e., .74 vs. .73), but accuracy was higher in the second block than in the first block for both RP+ (i.e., .80 vs. .85, $t = 3.0$) and RP−(i.e., .62 vs. .73, $t = 4.7$) items. The difference in accuracy between NP and RP− items was larger in the first block than in the second block (i.e., .13 vs. .00, $t = 4.4$), and only the first block difference was significantly different from zero ($t = 6.2$).

The measure of memory control was the difference in accuracy between NP and RP− items. The mean difference was small (i.e., .07, with a standard deviation of .26), but it was significantly greater than zero ($t = 4.6$). The estimated reliability of the NP-minus-RP− difference based on the values in the two trial blocks was .05. The correlation of age with the NP-minus-RP− difference was .09, and the semi-partial correlation between age and RP− accuracy after control of the variation in NP accuracy was −.13. Retrieval practice was associated with enhanced memory as the RP+-minus-NP difference was significantly greater than zero (i.e., mean = .09, standard deviation = .21, $t = 7.6$), but the amount of enhancement was not related to age (i.e., $r = -.01$).

Validity of a memory control construct

Table 5 contains information relevant to the computation of an estimate of the reliability of memory control difference scores from the reliabilities of the component scores and the correlations between them, based on the formula in the bottom of the table. As with the direct reliability values mentioned above, all of the computed estimates were quite low. Inspection of the table suggests that two different factors contributed to the low reliabilities. In several tasks, such as Multiple Trial Word Recall, Brown–Peterson, and Episodic Retrieval Inhibition, the reliabilities of the component scores were in the moderate range, but the correlations between the measures were nearly the same magnitude as the reliabilities. Because this indicates that the measures share nearly all of their reliable variance, as reflected in the finding that the estimates of the reliability-adjusted correlations were very close to 1.0, there was little or no variance associated with the difference between the measures. The correlations between the component measures were lower in the other tasks, but the reliabilities of one or both com-

Table 5
Reliabilities of derived memory control scores

Task	A	B	r_{AA}	r_{BB}	r_{AB}	A–B rel.		r_{AB*}
						Est.	Direct	
Multiple trial word recall	Trial 4	Trial 5	.72[a]	.72[a]	.72	0	NA	1.00
Picture directed forgetting	Remember	Forget	.49	.38	.09	.38	.41	.21
Word directed forgetting	Remember	Forget	.53	.17	.16	.23	.46	.53
Brown–Peterson proactive interference	Trial 1 & 4	Trial 3 & 6	.62	.60	.54	.15	.20	.89
Color word proactive interference	Trial 1	Trial 4	.17[b]	.31[c]	.08	.17	NA	.35
Semantic retrieval inhibition	NP	RP–	.62	.34	.52	−.08	−.10	1.13
Episodic retrieval inhibition	NP	RP–	.56	.66	.66	−.15	.05	1.09

Note. The estimated reliability of the A−B difference (Est.) was computed from the following formula: $r_{(A-B)(A-B)} = ([(r_{AA} + r_{BB})/2] - r_{AB})/(1 - r_{AB})$ from Cohen and Cohen (1983, p. 69) and the disattenuated correlation (r_{AB*}) was computed from the following formula: $r_{(AB^*)} = r_{AB}/\sqrt{(r_{AA} * r_{BB})}$ from Cohen and Cohen (1983, p. 69). Direct refers to the reliability of the difference obtained from coefficient alpha with difference scores from different parts of the test used as "items." NA means that a direct computation of reliability was not available. The superscripts in the table have the following meanings.

[a] The value was estimated from the median correlation across trials 1 and 2, 2 and 3, and 3 and 4.

[b] The value was estimated from the correlation between trials 1 and 2.

[c] Value was estimated from the correlation between trials 3 and 4.

ponent measures also tended to be low, which contributed to low estimated reliability of the difference score.

Correlations among the memory control measures are reported in Table 6. As one would expect from the low reliabilities of many of the individual memory control measures, the correlations among the memory control measures were all very weak. Only two correlations were significantly different from zero, and the correlation between the multiple trial word recall and word directed forgetting measures was very small (i.e., −.15). The significant correlation between the R–F difference scores for pictures and for words is not very interesting because it is largely attributable to the relations between the R measures. Indeed, the correlation of the F measures with one another after controlling the variation in the R measures was only .26, whereas the correlation of the R measures with one another after controlling the variation in the F measures was .55.

To demonstrate convergent validity the measures should have moderate to large correlations with one another. Because there was no evidence of relations among the measures, or that the measures had any variance in common, the results summarized in Table 6 provide no support for the existence of convergent validity for one or more memory control constructs based on these variables. Discriminant validity is determined by relations of the construct with constructs representing other cognitive abilities. However, it is not meaningful to examine discriminant validity when there is no evidence that the relevant variables have any variance in common, and might represent a coherent construct.

Influences of age and cognitive abilities on individual variables

The model in Fig. 2 was applied with each of the measures derived from the experimental tasks serving as the target variable with the AMOS (Arbuckle, 2003) structural equation program. As noted in the introduction, coefficients for the relations of the established cognitive abilities to the target variable are informative

Table 6
Means and correlations of memory control measures

	Mean	*SD*	1	2	3	4	5	6	7
1—Pre–post interfer.	2.0	1.8	(.00)						
2—Pict R–F	3.0	3.2	−.08	(.38)					
3—Word R–F	3.3	2.5	−.15*	.46*	(.23)				
4—Brown–Peterson PI	1.9	1.3	.07	.02	.08	(.15)			
5—Color PI	.18	0.43	−.05	−.06	.05	.09	(.17)		
6—Semantic RI	−.01	0.16	−.02	.10	.00	.08	−.05	(−.08)	
7—Episodic RI	.07	0.26	.04	.04	−.09	.03	−.02	.08	(−.15)

Note. Values in parentheses along the diagonal are reliabilities estimated by the formula in Table 5.

* $p < .01$.

about what the variable represents in terms of various cognitive abilities, and coefficients for the direct relations from age indicate the extent to which the age-related effects on the variable are independent of age-related effects operating through the reference abilities. Both sets of coefficients, in standardized units, for all of the measures from the six experimental tasks are presented in Table 7. The second column of the table contains the simple correlation of the target variable with age, and the third column contains the direct relation of age on the variable after controlling age-related influences through the cognitive abilities. The remaining columns in the table contain the standardized coefficients from each cognitive ability to the variable.

Table 7
Standardized coefficients of age and cognitive abilities on individual variables

Variable	Only age	Unique age	Cognitive ability			
			gF	Mem	Speed	Voc
Multiple trial word recall						
Trial 1	−.29*	.15	−.04	1.00*	.00	−.16
Trial 2	−.30*	.04	−.02	.97*	−.04	−.01
Trial 3	−.30*	.07	−.04	.97*	−.00	−.09
Trial 4	−.33*	.05	−.01	.91*	.02	−.05
List B	−.27*	.22	.02	.77*	.15	−.14
Trial 5	−.31*	.02	−.14	.89*	−.00	−.06
4–5	.12	.11	.15	−.29*	.05	−.05
Directed forgetting						
Words—R	−.39*	−.20	−.09	.42*	.14	.17
Words—F	−.10	−.03	−.04	.22	−.03	−.01
Pictures—R	−.29*	.17	.07	.48*	.27*	−.14
Pictures—F	.03	.07	−.45*	.44*	.10	.02
Words R–F	−.33*	−.17	−.07	.31*	.15	.17
Pictures R–F	−.27*	.11	.33*	.16	.17	−.14
Brown–Peterson						
Subtractions	−.13	.25*	.41*	.06	.29*	.14
Avg. list 1 & 4	−.31*	−.02	.03	.50*	.14	.14
Avg. list 2 & 5	−.22*	−.04	−.01	.52*	−.00	.16
Avg. list 3 & 6	−.27*	.14	.31*	.42*	.04	−.17
(1&4)–(3&6)	−.01	−.18	−.32*	−.01	.09	.32*
Color word						
Word Recall	−.43*	−.01	.14	.63*	.10	−.07
List 1 Recog.	−.08	−.08	−.11	.29*	−.03	.15
List 2 recog.	−.10	−.07	.05	.23	−.12	.01
List 3 recog.	−.12	−.06	−.04	.15	.00	−.06
List 4 recog.	−.06	−.15	−.11	.07	−.12	−.04
1–4	.02	.08	.03	.12	.06	.13
Semantic retrieval inhibition						
Integrat. freq.	−.19*	−.02	−.05	.07	.27	.05
Practice RP+	−.18*	−.17	−.01	.34*	−.12	.22
RP+	−.17*	−.08	.01	.34*	−.03	.18
RP−	−.05	.05	−.02	.19	.07	.06
NP	−.09	.03	−.08	.41*	.02	.15
NP–RP−	−.02	−.03	−.05	.13	−.05	.06
Episodic retrieval inhibition						
Trials to criterion	.30*	.12	−.15	−.32*	.02	−.08
Practice RP+	−.35*	−.16	.04	.57*	−.06	.05
RP+	−.36*	−.14	.15	.47*	−.06	−.02
RP−	−.30*	−.10	.07	.46*	.01	.03
NP	−.28*	−.12	−.01	.49*	.01	.10
NP–RP−	.09	.00	−.10	−.07	.01	.07

* $p < .01$.

All of the recall measures in the multiple trial word recall task were strongly related to memory ability, which is not surprising because the sum of the number of words recalled across trials 1 through 4 served as one of the indicators of episodic memory ability. The weakest relation from memory ability was to the measure of the difference between trial 4 and trial 5, and it was negative, indicating that people with higher memory ability had a smaller decrease in recall after the presentation and recall of a different list of words. None of the relations with the other cognitive abilities were significant for any of the measures in this task.

The recall measures in the directed forgetting task had moderate relations with memory ability, and the relations were slightly less positive on the measure of recall of items that were to be forgotten than on the recall of items that were to be remembered. However, it is worth noting that the relation between recall of to-be-forgotten items and memory ability might have been expected to be negative if effectiveness at discarding or inhibiting items after encoding was associated with better memory. Instead the relations were positive, indicating that people who had higher levels of episodic memory ability remembered more of the to-be-forgotten items than people with lower episodic memory ability.

The measure of rate of subtractions in the proactive interference task was positively related to gF ability and to perceptual speed, but was not related to memory ability. There were similar moderate relations of memory ability on the recall measures for each trial in this task, but the difference measure representing proactive interference was related to the gF and vocabulary abilities. Individuals with higher levels of gF had smaller proactive interference, but individuals with higher levels of vocabulary had greater proactive interference. The word recall measure in the color word task had a strong relation to memory ability, but the relations of the cognitive abilities on the color recognition measures were weak, and only one was significantly different from zero.

Many of the measures from the semantic and episodic retrieval inhibition tasks were moderately related to memory ability. In the semantic version of the task the relation of memory ability to RP− accuracy was somewhat weaker than that with the NP accuracy measure, but this was not the case to the same degree in the episodic version of the task. There were no relations of the cognitive abilities to the reported frequency of integrating the items at study in the semantic retrieval inhibition task. The trials-to-criterion measure in the episodic retrieval inhibition task was lower when memory ability was high, indicating that learning was faster among people with better memory abilities. Finally, the NP-minus-RP− difference was not significantly related to any of the cognitive abilities in either the semantic or episodic version of the task.

Two points regarding the results summarized in Table 7 should be emphasized. First, only three of the candidate measures of memory control had significant influences of the cognitive abilities, and one was not particularly interesting because it represented relations with episodic memory. The two variables with significant relations to gF were attributable to individuals with higher levels of fluid ability recalling fewer to-be-forgotten pictures, and hence exhibiting larger picture R-F scores, and recalling more words on the third list with the same semantic category, and thus exhibiting smaller PI effects. Of course, the lack of significant relations is not particularly surprising given the low reliability of the measures (cf. Table 5).

And second, the majority of the memory measures from the experimental tasks had significant correlations with age, but none of them had unique age relations that were independent of the age-related influences operating through the cognitive abilities. The only significant direct relation of age was with the number of subtractions measure in the proactive interference task, and it was positive, indicating that after taking into consideration age-related influences through the cognitive abilities, increased age was associated with a higher rate of subtractions.

Discussion

Many of the findings from previous research with the tasks selected to assess memory control were replicated in this study. To illustrate, recall of to-be-remembered items was significantly higher than recall of to-be-forgotten items in the directed forgetting task, recall performance was significantly lower after presentation and recall of a different list of words, and recall was progressively lower across successive lists of words from the same semantic categories. However, the expected difference in accuracy between NP and RP-items in both the semantic and episodic retrieval inhibition tasks was only significant in the first half of the trials. The failure to fully replicate the retrieval inhibition phenomenon is puzzling because the procedure in the semantic version of the task was nearly identical to the procedure in Anderson et al. (1994), and the level of RP+ accuracy in this study was very similar to the levels in the experiments in that report (i.e., mean of .83 compared to .81 in Exp. 1, .91 in Exp. 2, and .80 and .78 in Exp. 3). The episodic retrieval inhibition task was modified to make it easier than the version used by Ciranni and Shimamura (1999) because preliminary research revealed that many young adults found it difficult to learn the associations. Nonetheless, the levels of accuracy for RP+ items were still comparable across the studies as the average for adults between 18 and 39 years of age in the current study was .92, and the means were.95 and.88 for the college student participants in the corresponding conditions in Experiments 1 and 2 in Ciranni and Shimamura (1999). Unlike the other studies, however, overall accuracies of the NP and RP-items in this study were only slightly less than that of the RP+ items. Retrieval practice was associated with enhanced recall of RP+ items compared to NP items in both the semantic and episodic retrieval inhibition tasks, but in neither task was overall recall of RP-items significantly lower than overall recall of NP items, and thus there was no evidence of retrieval inhibition.

Although the model in Fig. 1 could not be applied to the investigation of a memory control construct because few of the memory control variables were measured reliably, it is still meaningful to apply the model in Fig. 2 to all of the measures from the experimental tasks. The major findings from these analyses were that most of the experimental measures were closely related to episodic memory ability, and that the only unique, in the sense of statistically independent, age-related influence after considering influences through the available cognitive abilities was on the subtraction rate variable. These results imply that only the subtraction rate measure from the experimental tasks requires a unique explanation for the age-related influences beyond that needed to account for the age-related influences on the cognitive abilities.

It is apparent in Figs. 4 through 7 that there were large age-related differences in many of the measures, but very few interactions. This pattern suggests that most of the influences associated with aging on the measures of performance in these tasks are not moderated by the particular conditions that were manipulated. The failure to find many unique age-related influences with the model portrayed in Fig. 2 also suggests that most of the age-related influences on the variables were shared with influences on the cognitive abilities. Although it is tempting to consider that the variable in which one is particularly interested represents something distinct, the model illustrated in Fig. 2 is one of the few analytical methods available for determining whether the age-related influences on a variable are statistically independent of age-related influences on other variables. The absence of interactions, and the discovery that many of the age-related effects on the memory measures are shared with variation in the cognitive abilities, therefore raises the possibility that age-related effects in these measures are more systemic than specific. Unfortunately, the nature of any systemic effect that might exist is not obvious, except that it does not appear to be related to the effectiveness of memory control as assessed with the measures examined in this study. Nevertheless, an intriguing implication of this perspective is that, as has been suggested for at least 15 years (e.g., Salthouse, 1992), the null hypothesis in research on aging and cognition should no longer be that there are no age differences, but rather that there are no age differences beyond those shared with other cognitive abilities. The reasons for the relations among cognitive variables and abilities are still not well understood, but they are very well-established, as are the age-related influences on them. Instead of ignoring these relations, therefore, future research should take them into consideration when investigating influences of age on new variables. The model portrayed in Fig. 2 is one way in which this can be done, but other methods of determining whether the age relations under investigation are merely another manifestation of a broader phenomenon should also be examined.

It should be emphasized that the results of these analyses indicate that there is statistical overlap of the age-related influences on the cognitive abilities and the target variables, but they do not indicate the nature of these influences, nor do they specify which variables or constructs are most fundamental. Cognitive ability factors were considered the "context" in these analyses primarily because they are well-established empirically and have good measurement properties. However, analogous types of analyses could have been conducted in which a set of process-based constructs served as the context if three or more reliable measures of each relevant construct could be identified. Efforts to extend the analyses in this direction should be pursued because the analytical procedure provides an objective method of determining the contribution of different factors on a particular variable. That is, a variable can be inferred to be influenced by a factor to the extent to which people who have high or low values on that factor differ in their level of performance on the target variable. Until more direct methods of quantifying the influence of various factors on a variable are developed, basing the determinations on comparisons of people who differ in the levels of established factors may be the most objective method available for investigating the "meaning" of a variable. It is clearly possible that different outcomes might be obtained if other constructs were to serve as the context in these types of analyses. Nevertheless, the discovery that most of the experimental variables were only related to episodic memory ability suggests that, at least from an individual differences perspective, the variables may not reflect much more than episodic memory as that construct is assessed by paired associates, word recall, and story memory tasks.

Although most of the effects that could be postulated to reflect the operation of memory control were significantly different from zero, and generally replicated the group-level effects from prior studies, very few of the memory control measures were reliable at the level of individual differences. As noted in the introduction, effects that are robust and replicable in experimental studies are not always informative in individual differences research if the relevant measures have very low reliability. That is, replicability and reliability are often inversely related in that between-person variation in the magnitude of an effect is "bad" when the significance of an effect is being evaluated, but it is "good" when attempting to identify correlates of the effect from an individual differences perspective.

Cognitive psychologists are often interested in what is assumed to differ between two variables because the difference is postulated to reflect a theoretically relevant process. However, as observed in this study, difference scores frequently have low reliability, and hence may not be useful in individual differences research. Contrary to what is sometimes assumed, this is not because difference scores are intrinsically unreliable (e.g., Rogosa & Willett, 1983), but instead it occurs because the original variables are often so highly correlated with one another that there is little variability in the difference.

The information in Table 5 can be used to illustrate the problem with the memory control variables in the current study. All of the memory control variables were based on a contrast between two variables, which can be designated A and B. As indicated in the bottom of the table, the reliability of the A-B difference is a function of the reliability of the component variables (i.e., r_{AA} and r_{BB}), and of the correlation between the two variables (i.e., r_{AB}). There are therefore two obvious possibilities for increasing the reliability of difference scores; increase the reliability of each component variable, or decrease the correlation between the two variables. The first option is fairly easy to accomplish because increasing the number of trials used to assess each variable would likely increase reliability of the component scores. The reliabilities of the component variables in Table 5 are not very high, and they could probably be increased with additional trials. However, this method may only be effective at increasing the reliability of difference scores when the correlation between the component variables is considerably smaller than the average reliability of the component variables. In the current study only the directed forgetting and color-word proactive interference contrasts are therefore promising candidates for increasing reliability by the addition of more trials.

The second method that could be used to increase the reliability of a difference score is to decrease the correlation between the component variables. That is, when the measures that are contrasted to yield an estimate of the effect share nearly all of their reliable variance, the measures may be functionally equivalent with respect to individual differences regardless of the mean levels of performance. Indeed, the last column of Table 5 indicates that after adjusting for reliability some of the correlations are very close to 1.0. This particular problem will not be solved by increasing the number of trials in each component task, and instead may require increasing the salience of, and hence the individual differences in, the critical processes that distinguish the variables being contrasted. In other words, the tasks will likely need to be modified to make the critical processes more prominent relative to processes that are common across the component variables. Some type of task redesign will therefore probably be required to increase the reliability of memory control tasks in which the component variables are highly correlated, such as the Multiple trial word recall, Brown–Peterson proactive interference, and semantic and episodic retrieval inhibition tasks in the current study.

Because the phenomenon of memory control could not be reliably assessed at the level of individuals, it is still an open question whether effectiveness of memory control decreases with age, and whether memory control measures are correlated with one another and might reflect a unitary ability. The challenge for researchers interested in these concepts is to find ways to increase reliability because until the critical measures can be assessed with at least moderate reliability, it will be impossible to interpret low correlations (or small group differences) involving these measures.

Several attempts to deal with the problem of weak reliability of difference scores are unsatisfactory. For example, including both variables in the analysis will result in the common variance dominating any unique variance, which defeats the purpose of focusing on what differs between the variables. Inclusion of only the variable hypothesized to contain the critical process is also not an effective solution because the variance common to both variables will be large whenever the correlation is large, and it would still overwhelm any unique variance even if only one of the variables were to be included in the analysis.

Use of residuals instead of differences only solves the problem that difference scores are not independent of the original variables (Cohen & Cohen, 1983), and it does not necessarily improve reliability because the residuals will also have little variance when the relevant variables are highly correlated. To illustrate, estimates of reliability were computed for each of the regression-based residuals of the memory control variables that were originally expressed as difference scores. The estimated reliabilities of the residuals were similar in magnitude to the reliability estimates for the difference scores (cf. Table 5) as they were .15 for the remember–forget contrast for words, .35 for the remember–forget contrast for pictures, .38 for the first versus third list contrast for PI, .05 for semantic retrieval inhibition, and .32 for episodic retrieval inhibition.

The rationale for a statistical approach to improving the quality of measurement is portrayed in Fig. 8. The top two panels illustrate measurement assumptions that are implicit in much contemporary cognitive research. The assumption of construct purity is that there is a one-to-one correspondence between a theoretical con-

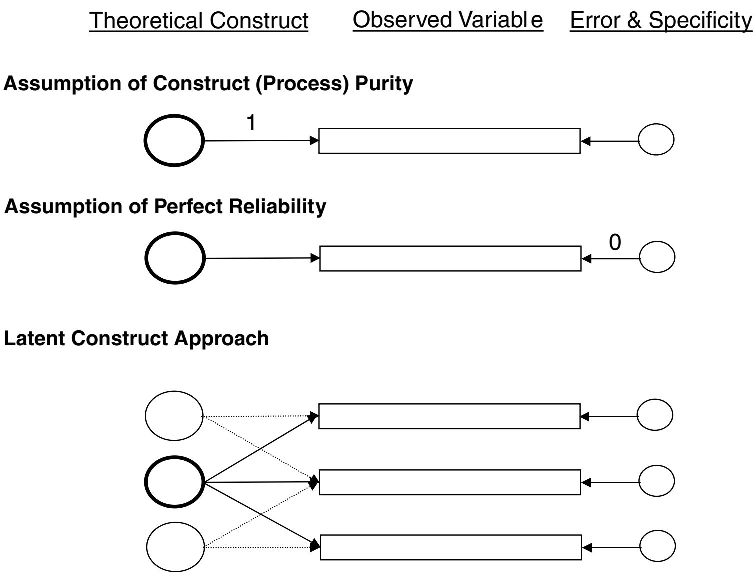

Fig. 8. Schematic representation of common measurement assumptions and the latent construct approach to assessment.

struct and an observed variable, and some version of this assumption seems to be operating whenever researchers rely on a variable from a single task to assess the theoretical construct of interest. The problem with this assumption is that most variables are probably not exclusively influenced by a single construct, and few constructs are likely to be exhaustively represented by a single variable.

The assumption of perfect reliability is that all of the variance in the measured variable is meaningful, and available to be associated with other variables. This assumption would probably not be explicitly endorsed by many researchers, but the general neglect of reliability in cognitive research suggests that it is often implicitly accepted. The problem is that few behavioral variables have no measurement error, either because of imprecision in assessment, or because of fluctuation in the relevant processes within the individual.

Neither the assumption of construct purity nor the assumption of perfect reliability is realistic, and acting as if they were valid can lead to theoretical and methodological problems such as those revealed in this study in which theoretically interesting variables were found to be of little value in individual differences research. An alternative approach to measurement that does not require making these assumptions, and instead provides estimates of the strengths of the critical relations, involves the use of multiple indicators to assess relevant theoretical constructs at a latent level. That is, careful selection of variables (cf. Little, Lindenberger, & Nesselroade, 1999) provides a more comprehensive representation of the construct than what is possible from any single variable, and a focus on common variance will minimize task-specific influences and emphasize shared influences. Furthermore, because only reliable variance can be shared, the common variance is theoretically free of measurement error. Structural equation models with latent constructs can also be used to specify latent difference scores that at least theoretically contain no measurement error (Donaldson, 1983; McArdle & Nesselroade, 1994), which will likely improve assessment of processes assumed to be indexed by a contrast between observed variables. The primary limitations of this approach are that the research is time consuming and expensive because multiple variables are needed for each construct, and the sample sizes must be relatively large to obtain precise estimates of the relations among variables.

The current study has many strengths as an investigation of individual differences in a construct like memory control. First, the moderately large sample size provides relatively high power and precision in estimating the relations among variables. Second, because the participants were not selected from an intellectually elite student population, or from the extremes of the distribution on one particular variable, sampling selection artifacts are unlikely. Third, the relevant constructs are assessed with multiple indicators to emphasize what they have in common, rather than relying on a single variable that likely only represents a limited aspect of the construct of interest, and is almost certainly influenced by other constructs. And fourth, influences on many cognitive abilities are examined simultaneously instead of focusing on a single ability, which can tempt researchers into concluding that the individuals differ only on the particular dimension that they have selected.

That is, merely because a researcher chooses to focus on variable X (e.g., working memory, executive functioning, etc.), he or she cannot be certain that the individuals are not also different on variables Y and Z unless these variables are included in the analysis.

In summary, there are four major findings of this study. First, many of the results from different paradigms that have been interpreted as reflecting the operation of memory control processes were replicated. Second, although the memory control effects were robust at the group level, measures of those effects at the level of the individual were not reliable, which precludes meaningful analyses of the interrelations among the variables. Third, examination of the relative influence of different cognitive abilities revealed that most of the experimental variables were closely related to episodic memory, and very few had influences of other cognitive abilities. And fourth, as in numerous other studies, age-related differences were found in a wide variety of cognitive variables but very few of those differences were unique, and statistically independent of the age-related influences on other variables.

Acknowledgments

We acknowledge the assistance of the following people in scheduling and testing participants and entering and checking data: Gauri Bhuchar, Kristina Caudle, Katherine Kane, Josh Magee, Cristina Rabaglia, Jena Saporito, Julia Siegel, Shannan Smith, and Lauren Walker. In addition, we thank James Miles for his help in programming several of the experimental tasks. This research was supported by NIA RO1AG19627 grant to T.A.S.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415–445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 1063–1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 608–629.

Arbuckle, J. L. (2003). *AMOS for windows. (Analysis of moment structures.)*. Chicago, IL: SmallWaters Corporation.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1997). *Differential aptitude test*. San Antonio, TX: The Psychological Corporation.

Bjork, R. A. (1998). Intentional forgetting in perspective; Comments, conjectures, and some directed remembering. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 453–481). Mahwah, NJ: Lawrence Erlbaum.

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

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 1403–1414.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Craik, F. I. M., & Salthouse, T. A. (2000). *Handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Donaldson, G. (1983). Confirmatory factor analysis models of information processing stages: An alternative to difference scores. *Psychological Bulletin, 94*, 143–151.

Dulaney, C. L., Marks, W., & Link, K. E. (2004). Aging and directed forgetting: Pre-cue encoding and post-cue rehearsal effects. *Experimental Aging Research, 30*, 95–112.

Dunlosky, J., & Salthouse, T. A. (1996). A decomposition of age-related differences in multi-trial free recall. *Aging, Neuropsychology, and Cognition, 3*, 2–14.

Earles, J. L., & Kersten, A. W. (2002). Directed forgetting of actions by younger and older adults. *Psychonomic Bulletin and Review, 9*, 383–388.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 145–199). Amsterdam: Elsevier.

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

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. (1982). The Toronto word pool: Norms for imageability, concreteness, orthographic variables, and grammatical usage for 1080 words. *Behavior Research Methods and Instrumentation, 14*, 375–399.

Gamboz, N., & Russo, R. (2002). Evidence for age-related equivalence in the directed forgetting paradigm. *Brain and Cognition, 48*, 366–371.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336–358.

Kausler, D. H. (1994). *Learning and Memory in Normal Aging*. San Diego: Academic Press.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 195–211.

McArdle, J. J., & Nesselroade, J. R. (1994). Using multivariate data to structure developmental change. In S. H. Cohen &

H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 223–267). Hillsdale, NJ: Lawrence Erlbaum.

Raven, J. (1962). *Advanced progressive matrices, Set II*. London: Lewis.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20*, 335–343.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Lawrence Erlbaum.

Salthouse, T. A. (1992). Shifting levels of analysis in the investigation of cognitive aging. *Human Development, 35*, 321–342.

Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences, 48*, P29–P36.

Salthouse, T. A. (2004a). What and when of cognitive aging. *Current Directions in Psychological Science, 13*, 140–144.

Salthouse, T. A. (2004b). Localizing age-related individual differences in a hierarchical structure. *Intelligence, 32*, 541–561.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532–545.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.

Salthouse, T. A., Berish, D. E., & Siedlecki, K. L. (2004). Construct validity and age sensitivity of prospective memory. *Memory & Cognition, 32*, 1133–1148.

Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology, 10*, 272–285.

Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (in press). Short-term variability and the calibration of change. *Journal of Gerontology: Psychological Sciences*.

Siedlecki, K. L., Salthouse, T. A., & Berish, D. E. (2005). Is there anything special about the aging of source memory? *Psychology and Aging, 20*, 19–32.

Wechsler, D. (1997a). *Wechsler adult intelligence ccale—3rd ed.*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler memory scale—3rd ed.*. San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock–Johnson psycho-educational battery—Revised*. Allen, TX: DLM.

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology, 89*, 109–116.

Zachary, R. A. (1986). *Shipley institute of living scale—Revised*. Los Angeles, CA: Western Psychological Services.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Lawrence Erlbaum.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 143–156.