

# Age Differences in the Monitoring of Learning: Cross-Sectional Evidence of Spared Resolution Across the Adult Life Span

Christopher Hertzog and Starlette M. Sinclair  
Georgia Institute of Technology

John Dunlosky  
Kent State University

Researchers of metacognitive development in adulthood have exclusively used extreme-age-groups designs. We used a full cross-sectional sample ( $N = 285$ , age range: 18–80) to evaluate how associative relatedness and encoding strategies influence judgments of learning (JOLs) in adulthood. Participants studied related and unrelated word pairs and made JOLs. After a cued-recall test, retrospective item strategy reports were collected. Results revealed developmental patterns not available from previous studies (e.g., a linear age-related increase in aggregate JOL resolution across the life span). They also demonstrated the value of investigating multiple cues' influences on JOLs. Multilevel regression models showed that both relatedness and effective strategy use positively and independently influenced JOLs. Furthermore, effective strategy use was responsible for higher resolution of JOLs for unrelated items (relative to related items). The effects of relatedness and strategy use with JOLs did not interact with age. The monitoring of learning is spared by adult development despite age differences in learning itself.

*Keywords:* learning, metacognition, strategy use, judgments of learning, aging

Researchers of metacognitive development in adulthood have uncovered an interesting age-related dissociation of monitoring associative learning and associative learning itself. Although associative learning (and subsequent memory) declines from maturity to older age (Hertzog & Dunlosky, 2004; Kausler, 1994; Naveh-Benjamin, 2000; Shing, Werkle-Bergner, Li, & Lindenberger, 2008), the ability to monitor learning seems largely unaffected by aging (Hertzog & Hultsch, 2000). The goal of this article is to critically evaluate the hypothesis that metacognitive monitoring of learning is maintained across the adult life span.

Monitoring of learning is often measured by asking individuals to rate their confidence that they can remember the item they just studied, which is called a *judgment of learning* (JOL). Paired-associate (PA) items are typically used for assessing metacognitive monitoring with JOLs because of their advantages for prompting both JOLs and recall with the cue alone (Nelson, 1996).

## Multiple Sources of Influence on JOLs

A critical issue is how individuals use multiple sources of information to construct JOLs (e.g., Dunlosky & Metcalfe, 2009;

Koriat, 1997). Current metacognitive theories emphasize that JOLs are based on access to information available during encoding or that can be retrieved from memory (Nelson, 1996). For example, Koriat's (1997) cue-utilization theory distinguished whether JOLs were based on what he called intrinsic, extrinsic, or mnemonic cues<sup>1</sup> (cf. Dunlosky & Matvey, 2001). According to Koriat's theory, intrinsic sources (which are specific to the stimuli themselves) have similar influences on JOLs and recall, whereas extrinsic sources (which are extrinsic to stimuli, such as how stimuli are processed) have a smaller influence on JOLs than on recall (i.e., people discount extrinsic sources when making JOLs).

The relative accuracy of JOLs, also known as resolution, is assessed by computing intraindividual (within-person) correlations of JOLs with PA recall outcomes (Nelson, 1984). These correlations indicate whether a person's JOLs covary with the probability of item recall. JOL resolution is influenced by whether individuals access diagnostic sources of information (i.e., sources that are correlated with the likelihood of later recall) when making a JOL. Observable stimulus characteristics, such as word frequency, concreteness, and the associative relatedness of elements of a PA item (intrinsic sources), are often related to the subsequent probability of recall. Therefore, use of these sources of information when making JOLs will enhance JOL resolution. Conversely, reliance on nondiagnostic sources can impair resolution.

Metacognitive illusions occur when a source of information influences JOLs to a different degree than it influences PA recall (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003; Rhodes &

---

Christopher Hertzog and Starlette M. Sinclair, School of Psychology, Georgia Institute of Technology; John Dunlosky, Department of Psychology, Kent State University.

This research was supported by National Institute on Aging Grant R37-AG13148 (Christopher Hertzog, principal investigator). We thank Daniela Jopp for coordinating the data collection, which was also assisted by Teri Boutot, Aaron Bozorg, Devaki Kumarhia, Shannon Langston, Colin Malone, Lulua Mandviwala, Melissa McDonald, David Winograd, and Helen Yu. More information on research in the Hertzog laboratory can be obtained at <http://psychology.gatech.edu/CHertzog/>

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332–0170. E-mail: [christopher.hertzog@psych.gatech.edu](mailto:christopher.hertzog@psych.gatech.edu)

---

<sup>1</sup> The term *cue* is often used to denote information used to make metacognitive judgments, as in the cue-utilization theory of Koriat (1997). To avoid semantic confusion, we reserve in this article the term *cue* to refer to the word from a PA item used to prompt a cued-recall trial, and we use *sources of information* or *sources* as the terms indicating the basis for making a JOL on a given trial.

Castel, 2008). For example, Koriat and Bjork (2005) demonstrated that people give associatively related items higher JOLs than unrelated items, but ignore asymmetries in forward versus backward cueing between the two words in a PA item (e.g., *cheddar* activates *cheese* as an associate but not vice versa). Because cued recall is better when there is forward association from cue to target, JOLs for backward associations are higher than their actual likelihood of recall. To account for such findings, Koriat and Bjork (2006) contrasted theory-based versus experience-based sources of influences, claiming that people's implicit theories or heuristics about the influence of stimulus characteristics could be misleading (as described above), but that these effects could be overcome by actual learning experience (see also Finn & Metcalfe, 2008; King, Zechmeister, & Shaughnessy, 1980).

Metacognitive illusions may be overcome when individuals weigh multiple sources of information (e.g., Koriat & Bjork, 2006). Accordingly, in the present study we evaluated two major sources of influence on JOLs and on age differences in JOLs—associative relatedness (an intrinsic source) and use of effective encoding strategies (an extrinsic source). Fluency of encoding affects metacognitive judgments at encoding, including JOLs (Benjamin, Bjork, & Schwartz, 1998; Hertzog et al., 2003; Robinson, Hertzog, & Dunlosky, 2006). One published study demonstrated that spontaneous use of effective strategies affects quality of encoding ratings (Dunlosky, Kubat-Silman, & Hertzog, 2003), but it is unknown whether people's JOLs are influenced by such strategy use. Mediators such as creating a sentence or an image to bind the new association improve PA learning (e.g., Dunlosky & Hertzog, 1998; Dunlosky, Hertzog, & Powell-Moman, 2005; Richardson, 1998). Hence, JOL resolution could be enhanced if JOLs are based on the quality of encoding strategies for different items. JOLs do correlate with reported success in implementing instructed strategy use (Robinson et al., 2006).

In our study, we evaluated the joint influences of associative relatedness and effective strategy use on JOLs. The two sources of information could operate independently to influence judgments, or their influences could be interrelated. Hertzog, Kidder, Powell-Moman, and Dunlosky (2002) demonstrated that resolution for a PA list containing related (e.g., *king–crown*) and unrelated (e.g., *turtle–bean*) items was higher than resolution for its subsets of unrelated and related items, showing that attending to associative relatedness benefitted JOL resolution. Moreover, Hertzog et al. (2002) found that resolution was better for unrelated items than for related items. The latter effect could be explained by the hypothesis that JOLs for related and unrelated items are equally influenced by whether an effective encoding strategy was used, but that use of effective meditational strategies has a larger effect on PA recall for an unrelated item (Dunlosky & Hertzog, 1998). That is, if JOLs for related items are influenced by use of normatively effective strategies, but strategies matter less for related items, then JOL resolution would be lower for related items. We tested that hypothesis in our experiment.

### Developmental Aspects of the Monitoring of Learning

Rabinowitz, Ackerman, Craik, and Hinchley (1982) showed that younger and older adults' JOLs for PA items were similarly affected by associative relatedness (see also Connor, Dunlosky, & Hertzog, 1997; Hertzog et al., 2002). Most age-comparative stud-

ies of JOLs have found equivalent resolution in younger and older adults (but see Daniels, Toth, & Hertzog, 2009).

Resolution of JOLs can be influenced by multiple variables (e.g., Dunlosky & Nelson, 1994; Finn & Metcalfe, 2008; Koriat, 1997). In virtually all cases we are aware of, variables that influence younger adults' JOL resolution for associative cued recall also influence older adults' JOL resolution in a similar manner (e.g., Dunlosky & Hertzog, 2000; Eakin & Hertzog, 2006; Hertzog et al., 2002; Robinson et al., 2006). The similarity of experimental effects supports the argument that processes of making JOLs are equivalent in younger and older adults.

### The Need for Full Cross-Sectional Data

To date, however, all of the studies of adult age differences in JOLs have employed extreme age-groups designs (Hertzog, 1996) that compare younger adults (usually, university students) with older adults, typically with a mean age in the early 70s. We know nothing about the developmental function relating the full range of adult ages to JOLs. Although one might expect from the existing evidence that the developmental function would be a relatively flat line, it could also be the case that there is quadratic curvature in the function, with a peak performance in middle age and a decline thereafter. Moreover, using university students as a reference group could be problematic, given that high-ability younger adults (selected on the basis of admission to college) are compared with a more heterogeneous group of older adults. Intellectual abilities predict episodic memory and associative learning (e.g., Hertzog, Dunlosky, & Robinson, 2009; Hultsch, Hertzog, Dixon, & Small, 1998; Kyllonen, Tirre, & Christal, 1991), as well as strategy use in associative learning (Hertzog et al., 2009). One can wonder whether the use of a select younger reference group biases estimated age effects in the monitoring of learning. Without full cross-sectional data, it is impossible to know.

We collected JOLs in a cross-sectional sample of adults, predicting cross-sectional age differences in PA recall favoring younger adults, but no age differences in the resolution of JOLs. We also hypothesized that (a) JOL resolution for unrelated items would be uniformly higher than the JOL resolution for related items across the adult life span; (b) the two sources of information, associative relatedness and use of effective encoding strategies, would impact JOLs and JOL resolution across the adult life span; but (c) effects of strategy use, the extrinsic cue (Koriat, 1997), would have less impact on JOLs than relatedness, the intrinsic cue.

## Method

### Participants

The sample consisted of 285 paid volunteers, ages 18–81, from the greater Atlanta, Georgia, metropolitan area. We excluded one age outlier—an 85-year-old participant—to avoid leverage on our regression analysis. The participants were part of a larger cross-sectional study of memory, memory beliefs, and related constructs. They were recruited from an existing adult volunteer data base or responded to television and print advertisements soliciting their participation. The age distribution was roughly uniform across the adult age span. About 11% of participants under the age of 25 were university students who received extra credit in their psychology

courses. Table 1 describes the sample on relevant demographic, affective, and cognitive variables, dividing the sample into younger (ages 18–39), middle-aged (ages 40–59), and older (ages 60–81) adults. Race was correlated with age (fewer of our older adults were African Americans); because controlling on this variable did not affect inferences about other variables, including age, we do not report on race effects in this article. Typical age differences in these variables emerged (e.g., negative age correlations with perceptual speed (Salthouse & Babcock's, 1991, Letter Comparison and Pattern Comparison tests) and the Community Epidemiological Screening for Depression (Radloff, 1977), but positive age correlations with the Shipley vocabulary test (Zachary, 1991).

### Experimental Task and Procedure

The PA learning task was part of the second session of a two-session study. A Visual Basic (Visual Studio [Version 6.0]; Microsoft Corporation, 1998) program controlled the task on a personal computer. It was the second PA task administered to study participants. The list contained 30 related and 30 unrelated PA items. Related items were selected from the University of South Florida free-association norms (Nelson, McEvoy, & Schreiber, 1998), avoiding the highest two associates of any cue; items had a mean forward association strength of 0.044.

Participants were informed about mnemonic strategies (e.g., interactive imagery) and were given a brief description of the strategies so they could provide item-level strategy reports (Dunlosky & Hertzog, 1998). The task presented the PA items in random order, each for 8 s. After participants studied each item, the cue remained on the screen, and individuals made a JOL, responding to "How confident are you that in about 10 minutes from now you will be able to recall the second word of the item when prompted with the first word?" Integer responses from 0% to 100% confidence were required.

Recall was prompted by presenting each cue. Correct recall was scored if the first three letters typed matched the first three letters of the target (which uniquely identified it). After recall, item pairs were shown in their original study order, and individuals reported what strategy, if any, they had used to help learn each item by

selecting one of six options: 1—rote repetition, 2—interactive imagery, 3—sentence generation, 4—some other strategy, 5—no strategy, 6—tried to use a strategy but ran out of time.

### Measures and Statistical Methods

We computed the mean JOL and mean PA recall for each individual, within cells of our independent variables (e.g., separately for related and unrelated items). Resolution was measured by computing ordinal Goodman–Kruskal gamma correlations of JOLs with binary recall outcomes (failure, success) for each person. Traditionally, gamma correlations have been used to assess resolution of metacognitive judgments (Nelson, 1984), even though there are some issues with interpreting them (e.g., Benjamin & Diaz, 2008). Given that most recent JOL studies use gamma correlations to measure resolution, we used them here to facilitate comparisons of our cross-sectional results to earlier extreme age-groups studies. One problem is that gamma correlations have high standard errors of estimate when marginal distributions of JOLs or recall are extreme (e.g., correct PA recall of 95% and higher; Hertzog et al., 2002). Gamma correlations can only be computed when there is variability in JOLs and recall, so there are missing data for some participants.

We used SAS PROC MIXED (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006; SAS Institute, 2000) to evaluate hierarchical polynomial regression models on age using restricted maximum likelihood estimation. Chronological age was centered, and then the first four powers of age (linear, quadratic, cubic, and quartic) were evaluated for all dependent variables. Given the large sample size and the within-subjects design, significance was evaluated at a criterion of  $\alpha = .01$ . In analyses that included the within-subjects factor of relatedness, interaction tests (e.g., Age  $\times$  Relatedness) were evaluated first, and nonsignificant terms were tested and eliminated in reverse order (quartic, cubic, quadratic, and linear; see Cohen, Cohen, West, & Aiken, 2003). In all the analyses we conducted, only the linear effects of chronological age were reliable at the .01 level, so we save space by only reporting results with linear age effects and interactions. Within-subjects experimental effects were estimated in PROC MIXED by specifying an unrestricted error covariance matrix.

Table 1  
*Participant Characteristics*

Characteristic	Younger ( <i>n</i> = 100)			Middle aged ( <i>n</i> = 92)			Older ( <i>n</i> = 92)		
	%	<i>M</i>	<i>SD</i>	%	<i>M</i>	<i>SD</i>	%	<i>M</i>	<i>SD</i>
Female	50			56			54		
Caucasian	40			56			80		
African American	41			39			15		
Hispanic	3			3			2		
Chronological age in years		29.62	6.78		49.63	6.18		68.57	5.51
CES-D total depression score <sup>a</sup>		14.40	9.44		11.77	10.27		6.66	6.52
Years of education		16.38	2.29		16.67	2.34		16.27	2.33
Shipley vocabulary		30.78	5.32		32.39	5.18		33.19	4.68
Pattern Comparison		19.22	3.74		16.32	4.15		14.93	3.46
Letter Comparison		11.58	2.51		9.61	2.04		8.79	2.10

*Note.* CES-D = Community Epidemiological Screening for Depression.

<sup>a</sup> Scores can range from 0 to 40; 16 is often considered a cutoff for possibly depressed.

Effect size computation is complicated for nonorthogonal repeated measures models with covariates, especially when conducted in PROC MIXED. We report pseudo- $R^2$  statistics as effect sizes for chronological age (the covariate), and report effect size for within-subjects independent variables like relatedness using an extension of Cohen's (1988)  $d$  statistic,  $d^*$ , which expresses cell mean or marginal mean differences as a function of the appropriate pooled error term,  $d^* = (M_1 - M_2) / \sqrt{\text{pooled variance estimate}}$ . It can be interpreted as the number of standard deviations separating the means in question. Cohen (1988) suggested benchmarks of 0.2, 0.5, and 0.8 for small, medium, and large effect sizes, respectively.

**Results**

**PA Recall and Mean JOLs**

Figure 1 plots the linear age trends in PA recall (Panel A) and mean JOLs (Panel B) separately as a function of associative relatedness. To facilitate comparisons of the two variables across the two panels, we show both variables in percentages: recall in percentage correct and JOLs in percentage confidence.

As expected, mixed model analysis revealed reliable negative age differences in percentage of correct PA recall,  $F(1, 283) = 39.05, p < .001, R^2 = .11$ . Relatedness also had a potent effect on

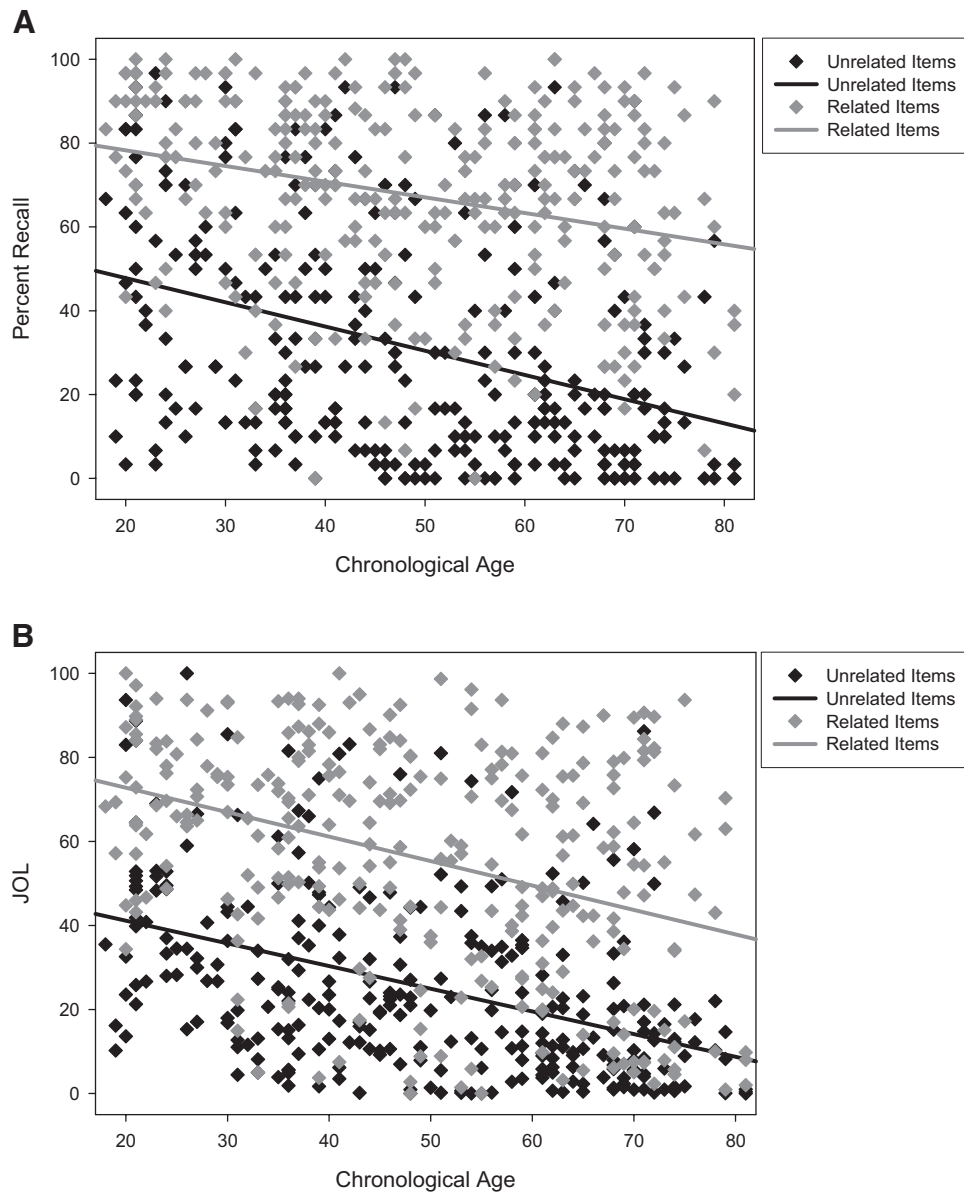


Figure 1. Scatterplots with overlaid regression lines of mean recall and judgments of learning (JOLs) for related and unrelated paired-associate items as a function of chronological age. A: Paired-associate recall. B: Mean JOLs.

PA recall,  $F(1, 283) = 1,013.68, p < .001, d^* = 1.51$ . The fitted least-squares means at the centered regression intercept (age 49) were 31 ( $SE = 2$ ) and 68 ( $SE = 1$ ) for unrelated and related items, respectively. This effect interacted with age, such that age differences over the adult life span were larger for unrelated than related items,  $F(1, 283) = 9.36, p < .01$ . The estimated slopes, when rescaled as change in percentage of items recalled per year of age, were  $-0.58$  ( $SE = 0.09$ ) and  $-0.37$  ( $SE = 0.08$ ), for recall of unrelated and related items, respectively. The  $R^2$  effect sizes were 13% and 8% of the sample variance, for unrelated and related items, respectively.

Like the PA recall data, participants' mean JOLs were strongly influenced by both age and relatedness. Reliable linear effects occurred for age,  $F(1, 283) = 63.52, p < .001, R^2 = .17$ , and relatedness,  $F(1, 283) = 660.49, p < .001, d^* = 1.34$ . The mean JOL for related items was about 30% higher than the mean JOL for unrelated items ( $M_{\text{Related}} = 56.1, SE = 1.5$  vs.  $M_{\text{Unrelated}} = 25.6, SE = 1.2$ ). Unlike PA recall, there was no hint of an interaction of relatedness with age on JOLs ( $F < 1$ ). Given the large sample size and small  $F$  statistic, it is reasonable to assume that the regression lines for related and unrelated items are essentially parallel. Indeed, the fitted slopes were very close in magnitude, being  $-0.58$  ( $SE = 0.09$ ) and  $-0.54$  ( $SE = 0.07$ ) for related and unrelated items, respectively. The  $R^2$ s for age effects were 14% and 18% of the variance in related and unrelated mean JOLs, respectively.

Comparing the two panels of Figure 1 indicates that subjective confidence was consistently lower, in the aggregate, than was PA recall across the adult life span, when both were plotted on the same percentage scale. JOLs had a steeper linear slope for related items than was the case for related PA recall, so there was also an increasing disparity between JOLs and PA recall with increasing age, indicating an age difference in absolute accuracy of JOLs for related items. This was not the case for unrelated items, where the two lines were essentially parallel.<sup>2</sup>

## Resolution

Figure 2 plots the linear age trends for the gamma correlations relating JOLs to PA recall, in the aggregate and separately as a function of associative relatedness. The aggregate gamma correlation, combining related and unrelated items, shows resolution when relatedness is used as information for making accurate JOLs. As can be seen in Figure 2, the aggregate gamma correlation was uniformly the highest of the three correlations.

The mixed model analysis revealed an unexpected, reliable increase in the aggregate gamma correlation across the adult age span,  $F(1, 279) = 10.77, p < .001, R^2 = .04$ . The estimated slope was 0.0035 ( $SE = 0.001$ ). The overall mean gamma correlation was .51; from age 20 to 80, the predicted correlation increased from .40 to .62.

When the data were separated into unrelated and related items, the gamma correlations showed the hypothesized main effect of relatedness, with higher gamma correlations for unrelated items,  $F(1, 278) = 12.45, p < .001, d^* = 0.29$ . On average, the mean gamma correlation was .28 ( $SE = .03$ ) for unrelated items and .15 ( $SE = .03$ ) for related items. Both values were reliably greater than zero, indicating above-chance resolution of JOLs for both item types. The overall effect of age was not reliable ( $F < 1$ ), nor was there a reliable Age  $\times$  Relatedness interaction,  $F(1, 278) = 2.80,$

$p > .05$ . Despite the apparent linear increase in gammas for unrelated items in Figure 2, the regression slope separately estimated for unrelated items also did not achieve statistical significance ( $b = 0.0024, SE = 0.0017, t = 1.43, p > .10$ ).<sup>3</sup>

## Mediator Strategy Use

We analyzed the outcomes of the item-level strategy reports by computing the aggregate likelihood of people using an effective mediational strategy, pooling reports of different effective strategies (imagery use, sentence generation, and reported other strategies; see Hertzog et al., 2009) into a single variable. A mixed model analysis revealed robust age differences,  $F(1, 283) = 9.30, p < .01$ , reflecting an age-related decrease in effective strategy use. There was also a large effect of relatedness,  $F(1, 283) = 276.05, p < .001$ . On average, producing effective strategies was more likely for related items ( $M = 0.78, SE = 0.02$ ) than for unrelated items ( $M = 0.49, SE = 0.02, d^* = 0.96$ ), possibly because it is easier to identify ways of relating normatively associated concepts. The Reliable Age  $\times$  Relatedness interaction,  $F(1, 283) = 45.84, p < .001$ , revealed that age differences in strategy production were greater for the more difficult unrelated items. There were reliable age differences in effective strategy use for unrelated items ( $r = -.32, p < .01$ ), but not for related items ( $r = .04$ ).

Analyses we do not report in detail here showed potent effects of effective strategy use on PA recall (mean recall was 25% for ineffective strategies compared to 64% for effective strategies;  $d^* = 1.88$ ) replicating previous research (e.g., Dunlosky & Hertzog, 1998). Hence effective strategy use could be a valid source of information for JOLs and could account for some of the aggregate gamma correlations of JOLs with recall. The gamma correlations were reliably greater than zero (chance resolution) when separated into ineffective ( $\gamma = .41, t(277) = 9.34, p < .001$ ), and effective strategies ( $\gamma = .26, t(277) = 10.90, p < .001$ ). Controlling on effective strategy use reduced resolution from the aggregate gamma correlation, indicating that strategy use was a cue that contributed to aggregate JOL resolution.

The difference between ineffective and effective strategies in JOL resolution shown above was statistically reliable,  $F(1, 273) = 8.85, p < .01, d^* = 0.28$ . The lower resolution for ineffective strategies could represent influences of idiosyncratic encoding on some items that affected both JOLs and recall. This effect did not interact with age ( $F < 1$ ).

## Joint Influences of Relatedness and Strategy Use on Recall, JOLs, and JOL Accuracy

**Recall.** We conducted PROC MIXED analyses of percentage PA recall using both effective strategy use and relatedness as

<sup>2</sup> These observations were consistent with a general linear model run on the difference scores between mean PA recall and mean JOLs that we do not report here.

<sup>3</sup> Part of the issue is that the standard error for the fitted gamma correlations was large, given the presence of a substantial number of  $-1.0$  gamma correlations across the life span, a phenomenon often attributable to unstable estimates with highly skewed marginal distributions of recall or JOLs (see Hertzog et al., 2002).

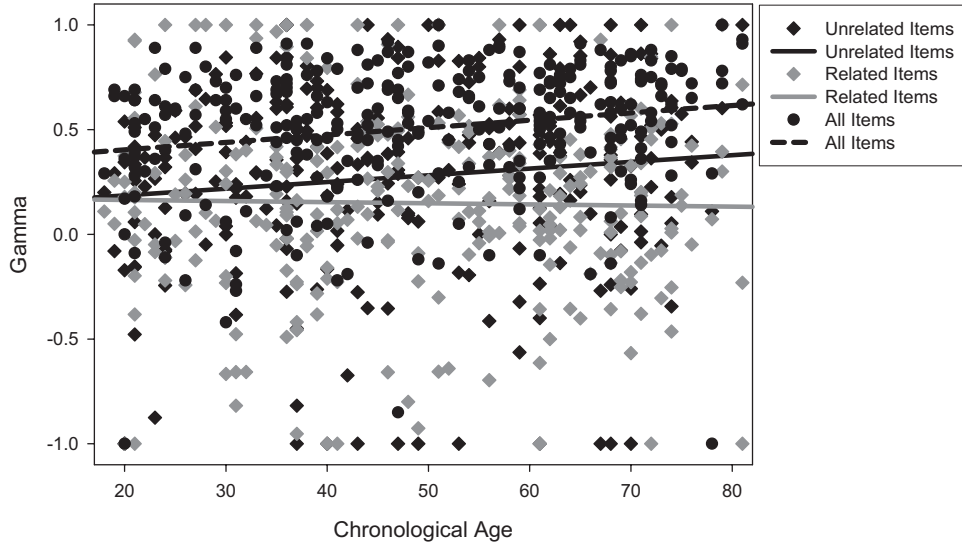


Figure 2. Scatterplots with overlaid regression lines for gamma correlations of judgments of learning with paired-associate recall, in the aggregate (all items) and separately for related and unrelated items.

independent variables, with chronological age as a continuous covariate. Effective strategy use,  $F(1, 283) = 428.29, p < .001, d^* = 1.32$ , and relatedness,  $F(1, 283) = 368.90, p < .001, d^* = 0.88$ , had large, independent effects on recall, as did age,  $F(1, 283) = 37.81, p < .001$ . Table 2 reports the fitted least-squares cell means and marginal means in the Relatedness  $\times$  Effective Strategy Use factorial. The adjusted age slope was  $-0.42 (SE = 0.09)$ , indicating a loss of just under half a percent recall per year of age. Unlike the analysis reported earlier, there was no Reliable Age  $\times$  Relatedness interaction when we controlled for effective strategy use ( $F < 1$ ), nor were there any other reliable age-related interactions. Apparently, the greater likelihood of effective strategy use for related items for older adults accounted for the Age  $\times$  Relatedness interaction seen in Figure 1A.

**JOLs.** We next evaluated whether JOLs were jointly and independently influenced by item relatedness and effective strategy use, using the same PROC MIXED analysis. The key question was whether we would see independent influences of both sources of information on JOLs when they were simultaneously included in the mixed model regression analysis. There were reliable effects of age ( $b = -0.63, SE = 0.09$ ),  $F(1, 283) = 61.47, p < .001$ , relatedness,  $F(1, 283) = 713.08, p < .001$ , and effective strategy use,  $F(1, 283) = 149.49, p < .001$ , on mean JOLs (see Table 2 for the marginal means). The effect size was large for relatedness ( $d^* = 1.15$ ) and medium for strategy use ( $d^* = 0.49$ ). There was no effective Strategy Use  $\times$  Relatedness interaction ( $F < 1$ ). Thus, both sources of information had independent influences on JOLs.

Table 2  
Fitted Cell and Marginal Mean JOLs, Paired-Associate Recall, and JOL–Recall Gammas as a Function of Effective Strategy Use and Relatedness, Covarying on Age

Variable	Related items		Unrelated items		Marginals	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Effective strategy						
JOL	58.6	1.5	31.6	1.4	45.1	1.3
Recall	71.2	1.5	49.6	1.7	60.4	1.4
$\gamma$	0.05	0.04	0.09	0.04	0.06	0.03
Ineffective strategy						
JOL	47.3	1.6	19.7	1.6	33.5	1.2
Recall	37.5	2.1	11.9	1.3	24.7	1.4
$\gamma$	0.17	0.05	0.11	0.05	0.14	0.04
Marginals						
JOL	52.9	1.4	25.7	1.1		
Recall	54.4	1.4	31.7	1.2		
$\gamma$	0.11	0.03	0.10	0.04		

Note. Both judgments of learning (JOLs) and recall are scaled as percentages to facilitate comparison of their means.

The analysis also revealed small interactions of age with both relatedness,  $F(1, 283) = 3.88, p < .05$ , and effective strategy use,  $F(1, 283) = 4.07, p < .05$ , in the direction of smaller age differences for ineffective strategies and for unrelated items. The three-way interaction was not reliable ( $F < 1$ ). The fitted age slopes involving the strategy use variable were  $-0.66$  for effective strategies and  $-0.53$  for ineffective strategies; the age slopes were  $-0.60$  for related items and  $-0.45$  for unrelated items. This difference in age slopes for related versus unrelated items was inconsistent with the pattern shown in Figure 1B when effective strategy use was not included in the analysis. This outcome suggested that the combination of (a) the tendency to give items studied with effective strategies a higher JOL and (b) the higher probability of successful strategy use for related items masked the age differences in slopes when relatedness was evaluated without simultaneously evaluating strategy use. The larger point, however, is that these small ordinal interactions did not suggest major age differences in attending to relatedness and effective strategy use as possible sources of information about recall when making JOLs. In sum, relatedness and effective strategy use had independent effects on JOLs, and neither variable accounted for age differences in mean JOLs.

**JOL-recall correlations.** Remember that resolution (a) was high when all items were analyzed, (b) was reduced when separating items either into related and unrelated items or into trials with effective and ineffective strategy use, and (c) was reliably higher for unrelated items than related items when ignoring strategy use. Our next question was whether the reliable effects of effective strategy use on JOLs and recall would account for the resolution differences between related and unrelated items. This question was addressed by a mixed model analysis blocking on both independent variables, evaluating whether doing so would eliminate relatedness differences in resolution. Moreover, if blocking on both variables reduced resolution to chance, then one could argue that strategy use and relatedness were the primary sources of information creating above-chance JOL resolution.

Resolution was analyzed in a  $2 \times 2$  within-subjects design (effective strategy use by relatedness; see Table 2), including age as a continuous predictor variable. The difference in gamma correlations for related and unrelated items disappeared when we controlled for effective strategy use ( $F < 1, d^* = .02$ ). Conversely, controlling on relatedness eliminated reliable differences in resolution between ineffective versus effective strategy use,  $F(1, 277) = 2.55, p > .10, d^* = 0.13$ . Simultaneously controlling on both cues reduced associations of JOLs with PA recall. However, the marginal mean  $\gamma$  for related and unrelated items was still reliably different from zero ( $p < .01$ ). This outcome suggested that other unmeasured sources of information that had validity for predicting recall also had some residual influence on JOLs. Nevertheless, most of the JOL resolution found in the aggregate data was accounted for by relatedness, effective strategy use, or other sources that correlated with these two variables.

**JOL calibration.** Another form of judgment accuracy involves absolute accuracy (metric deviations of JOL magnitudes from likelihood of recall). Calibration is one form of absolute accuracy, in which JOLs are ordered from low to high, and deviations of probability of recall from level of confidence (rescaled as subjective probability of success) are computed (see Dunlosky & Metcalfe, 2009, for an introduction). We used a quantitative calibration index (Lichtenstein, Fischhoff, & Phillips,

1982), computed separately for each of the  $2 \times 2$  cells of Relatedness  $\times$  Effective Strategy Use. The calibration index summed absolute values of the difference between JOLs and probability of correct recall across JOL bins, defined by treating the 0% confidence and 100% confidence extremes as separate values, and aggregated the remaining JOLs into bins (1–20 [midpoint of 10], 21–40 [midpoint of 30], 41–60 [midpoint of 50], 61–80 [midpoint of 70], and 81–99 [midpoint of 90]). Higher scores on the index indicate poorer calibration, so it can be thought of as an index of miscalibration.<sup>4</sup>

A mixed model analysis showed that calibration was dramatically affected by effective strategy use, with poorer calibration for items studied with effective strategies,  $F(1, 283) = 176.31, p < .001, d^* = 0.91$ . Mean miscalibration was 0.21 for items studied with effective strategies, compared to 0.09 for items studied with ineffective strategies. This effect was moderated by a Strategy  $\times$  Relatedness interaction,  $F(1, 283) = 20.81, p < .001$ . Calibration was worst for related items studied with effective strategies ( $M = 0.24$ ), and the difference in miscalibration for effective versus ineffective items was largest for related strategies (mean strategy difference of 0.16 for related items vs. 0.09 for unrelated items; difference in  $d^* = 0.53$ ). Finally, there was a trend for a linear effect of age on calibration,  $F(1, 283) = 4.32, p < .05$ , tending toward worse calibration for older adults; however, this trend was qualified by a reliable three-way interaction of Age  $\times$  Strategy  $\times$  Relatedness,  $F(1, 283) = 10.12, p < .01$ . To facilitate interpretation, we show in Figure 3 a bar chart dividing age into three groups, younger, middle-aged, and older (using the same cut points as in Table 1). Older adults tended to have better calibration except when related items were studied with effective strategies, the condition leading to the highest levels of PA recall, where that pattern was reversed. Age correlations with the calibration index were small and negative in the other three cells ( $r_s > -.21$ );  $r$  was  $.17$  for related items studied with effective strategies.

## Discussion

In the present study, we are the first researchers to examine JOLs and JOL accuracy in a full cross-sectional sample, which provided unprecedented methodological and statistical power to explore developmental trends in JOLs and JOL resolution. Concerning age trends in JOLs, our regression analysis (see Figure 1) revealed that JOL magnitudes decline consistently across the adult life span. The changes in JOL magnitude also paralleled declines in memory performance, suggesting that either (a) people are aware of subtle changes in memory that arise during aging, or (b) JOLs decline for different reasons, such as an age-graded expectation or implicit theory of decline that is disconnected from awareness of actual decline (see McDonald-Miszczak, Hertzog, & Hultsch, 1995). Concerning JOL resolution, our results are largely consistent with prior findings from extreme-age-groups studies. Whereas PA recall reliably declines with age, the resolution of JOLs does not. In fact, one surprising outcome of this study was the modest increase in aggregate JOL resolution across the adult

<sup>4</sup> We did not have sufficient density of JOLs to produce stable, full calibration curves across the entire range of JOLs (see Keren, 1991), given that these curves would have been needed separately for all four Strategy Use  $\times$  Relatedness cells.

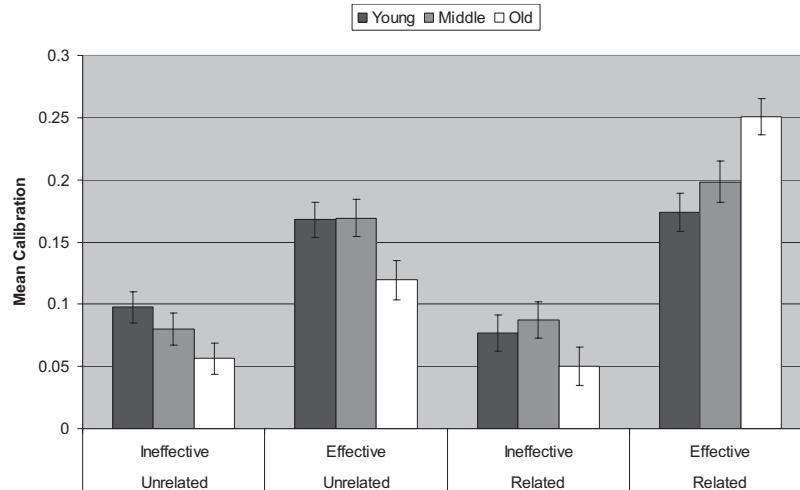


Figure 3. Calibration index scores for the four combinations of Relatedness  $\times$  Effective Strategy Use separately for younger, middle-aged, and older adults. Higher scores indicate poorer calibration. Error bars represent  $\pm 1 SE$ .

life span. Although this positive effect might be due to sampling error around a population regression coefficient of zero, the positive slope found in this cross-sectional sample makes it even less likely that the true population slope in adulthood is negative. We conclude that there are no age-related declines—and perhaps even improvements—in adults' ability to discriminate items they will recall from items they will not recall.

The magnitude of the aggregate gamma correlation of JOLs with PA recall was relatively substantial (about .50). One reason that it was well above chance was that individuals used and benefitted from the observable stimulus characteristic of associative relatedness. JOLs are influenced by relatedness, as is PA recall, and JOLs have greater resolution for the entire list than separately within the sets of related and unrelated items. The plotted regression curves show that adults of all ages use relatedness as a cue for JOLs.

We also found, as hypothesized, that use of effective encoding strategies influences JOLs and JOL resolution. These influences were statistically independent of associative relatedness. The retrospective strategy reports used in this study permit us to claim that spontaneous strategy use is a source of information influencing JOLs even when strategy use is not required by experimental instructions. These two variables (relatedness and strategy use) have been investigated only in isolation in previous studies, and the present evidence indicates that a separate evaluation of their influence leads to an incomplete understanding of JOL resolution. In particular, we replicated previous findings that resolution is greater for unrelated than related items (Hertzog et al., 2002). However, this difference was eliminated when strategy use and relatedness were both included in the analysis. The difference in resolution can therefore be attributed to the fact that JOLs are influenced by effective strategy use in both conditions, but that effective strategy use has a larger impact on PA recall for unrelated items than for related items (e.g., Dunlosky & Hertzog, 1998). In any case, individuals apparently attend to strategic processing outcomes when making JOLs (Robinson et al., 2006), even when they are also attending to the relatedness of the item pairs.

Our results indicate that a multiple-source approach to investigating metacognitive judgments is both fruitful and essential. Individuals apparently can and do attend to multiple variables (both intrinsic and extrinsic cues in Koriat's, 1997, terms) when constructing JOLs (see also Castel, 2008). As important, the statistical power of our design yielded the strongest evidence to date concerning Koriat's (1997) taxonomy of influences on JOLs. In the present data set, (a) both intrinsic and extrinsic sources influence JOLs and JOL resolution, and (b) both types of sources are discounted by JOLs in the limited sense that their absolute effect is larger on recall than on JOLs. However, the joint multilevel regression model showed that the effect size for relatedness on JOLs was larger than the effect size of relatedness on PA recall. Conversely, effective strategy use had a larger effect size on PA recall than it had on JOLs. Thus, the pattern of effect sizes indicates a larger impact of the intrinsic cue of relatedness than the extrinsic cue of using effective mediational strategies. Having said that, the robust effects of effective strategy use on JOLs contradict the hypothesis that individuals ignore or largely discount the extrinsic source of effective strategy use when making JOLs (cf. Shaughnessy, 1981).

We should also note that other cues besides relatedness and effective strategy use could have affected JOLs in this task—we simply did not measure them. Cues such as encoding fluency (Hertzog et al., 2003), font size (Rhodes & Castel, 2008), or serial order of presentation during study (Dunlosky & Matvey, 2001) have been shown to influence JOLs without also influencing JOL resolution because—unlike the variables we studied—they are not necessarily valid predictors of future recall.<sup>5</sup> In principle, the

<sup>5</sup> Encoding fluency can positively correlate with recall under some circumstances, such as when items vary in intrinsic cues like concreteness or associative relatedness. Hertzog et al. (2003) and Robinson et al. (2006) showed that with concrete, unrelated PA items, the fluency of generating interactive images under experimental instructions to do so correlated with JOLs but not with PA recall.

analytic approach we used here can be extended to study more than two observed or manipulated sources of information and how they affect item-level JOL magnitude and JOL resolution (see Hines, Touron, & Hertzog, 2009).

The multiple-source effects we found also provide further evidence of similar mechanisms of JOL construction across adulthood. For the most part, the influences of effective strategy use and relatedness on JOL resolution and JOL magnitude were well preserved across the adult life span. Age showed no sign of interacting with either independent variable in influencing JOL resolution. Age interacted with relatedness and effective strategy use in effects on JOLs, with smaller cross-sectional age differences on unrelated items and on items for which ineffective strategies were employed. However, these small ordinal differences in age slopes on JOLs for different item types and strategies suggests that people of all ages were influenced by relatedness and effective strategy used in a similar manner, even though, with increasing age, less differentiation was observed between either (a) related versus unrelated items or (b) items studied with ineffective versus effective strategies.

We have not emphasized age differences in absolute accuracy of the JOLs (as measured either by difference scores contrasting mean JOLs and mean recall or by a calibration index) because of concerns about the interpretability of absolute accuracy at different levels of age (see Connor et al., 1997). That is, given that we did not equate people of different ages on levels of memory performance, it is unclear exactly how to interpret age differences in absolute accuracy (Dunlosky & Metcalfe, 2009). The quantitative calibration index produced a complex pattern of age differences. Calibration was dramatically affected by whether an ineffective or effective strategy was used, suggesting that individuals did not vary JOLs sufficiently to reflect the large effect of producing integrated mediators on PA recall. Miscalibration was particularly large when effective mediators were produced for related items, and in this condition older adults had worse calibration than younger adults. Such effects could reflect a tendency for JOLs to be anchored near the midpoint of the confidence scale, failing to adequately calibrate to conditions leading to the highest or lowest recall levels (Connor et al., 1997; Scheck, Mateer, & Nelson, 2004). Further investigation of age differences in calibration in a study in which memory performance is equated across age levels would be needed to make definitive statements about age differences in absolute accuracy and calibration.

In summary, this study confirms that no age differences arise in the resolution of JOLs in a PA learning task, even when that task demonstrates age differences in PA learning itself. Although further explorations of the conditions under which age differences in JOL resolution can occur are warranted, such as when age differences in familiarity influence JOLs' resolution with recognition memory outcomes (e.g., Daniels et al., 2009), the present results speak against a general age deficit in the monitoring of encoding processes. Therefore, these data are encouraging about the validity of previous inferences about the sparing of monitoring made from extreme-age-groups designs. We caution that this kind of convergence may not always be found. Extreme-age-groups results should generally be evaluated with full cross-sectional evidence and later with longitudinal data, if possible (Hertzog, 1996), before one can confidently infer the existence of age-related effects over the adult life span.

## References

- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.
- Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York, NY: Psychology Press.
- Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, *36*, 429–437.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*, 50–71.
- Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and Aging*, *24*, 494–500.
- Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy use? *Psychology and Aging*, *13*, 597–607.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, *15*, 462–474.
- Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of five mediator-based deficiencies to age-related differences in associative learning. *Developmental Psychology*, *41*, 389–400.
- Dunlosky, J., Kubat-Silman, A., & Hertzog, C. (2003). Effects of aging on the magnitude and accuracy of quality-of-encoding judgments. *American Journal of Psychology*, *116*, 431–454.
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1180–1191.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565.
- Eakin, D. K., & Hertzog, C. (2006). Release from implicit interference in memory and metamemory: Older adults know that they can't let go. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *61*, P340–P347.
- Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*, 19–34.
- Hertzog, C. (1996). Research design in studies of aging and cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (4th ed., pp. 24–37). New York, NY: Academic Press.
- Hertzog, C., & Dunlosky, J. (2004). Aging, metacognition, and cognitive control. In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 215–251). San Diego, CA: Academic Press.
- Hertzog, C., Dunlosky, J., & Robinson, A. E. (2009). *Intellectual abilities and metacognitive beliefs influence spontaneous use of effective encoding strategies*. Unpublished manuscript.
- Hertzog, C., Dunlosky, J., Robinson, E., & Kidder, D. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22–34.
- Hertzog, C., & Hulstsch, D. F. (2000). Metacognition in adulthood and

- aging. In T. Salthouse & F. I. M. Craik (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Erlbaum.
- Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*, 209–225.
- Hines, J. C., Tournon, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: an analysis of adult age differences. *Psychology and Aging, 24*, 462–475.
- Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (1998). *Memory change in the aged*. New York, NY: Cambridge University Press.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.
- Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217–273.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing—The influence of retrieval practice. *American Journal of Psychology, 93*, 329–343.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187–194.
- Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1133–1145.
- Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology, 120*, 57–79.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute.
- McDonald-Miszczak, L., Hertzog, C., & Hultsch, D. F. (1995). Stability and accuracy of metamemory in adulthood and aging. *Psychology and Aging, 10*, 553–564.
- Microsoft Corporation. (1998). *Visual Basic* (Version 6.0) [Computer software]. Redmond, WA: Author.
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1170–1187.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from <http://web.usf.edu/FreeAssociation/>
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116.
- Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982). Aging and metamemory: The role of relatedness and imagery. *Journal of Gerontology, 37*, 688–695.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615–625.
- Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5*, 597–614.
- Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition, 13*, 458–478.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.
- SAS Institute. (2000). *SAS/STAT User's Guide* (Version 8). Cary, NC: Author.
- Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language, 51*, 71–79.
- Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20*, 216–230.
- Shing, Y. L., Werkle-Bergner, M., Li, S.-C., & Lindenberger, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General, 137*, 495–513.
- Zachary, R. A. (1991). *Shipley Institute of Living Scale—Revised Manual*. Los Angeles, CA: Western Psychological Services.

Received May 18, 2009

Revision received October 23, 2009

Accepted March 17, 2010 ■