

Age-Related Differences in Strategy Knowledge Updating: Blocked Testing Produces Greater Improvements in Metacognitive Accuracy for Younger than Older Adults

JODI PRICE¹, CHRISTOPHER HERTZOG¹, AND JOHN DUNLOSKY²

¹School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA, and ²Department of Psychology, Kent State University, Kent, OH, USA

ABSTRACT

Age-related differences in updating knowledge about strategy effectiveness after task experience have not been consistently found, perhaps because the magnitude of observed knowledge updating has been rather meager for both age groups. We examined whether creating homogeneous blocks of recall tests based on two strategies used at encoding (imagery and repetition) would enhance people's learning about strategy effects on recall. Younger and older adults demonstrated greater knowledge updating (as measured by questionnaire ratings of strategy effectiveness and by global judgments of performance) with blocked (versus random) testing. The benefit of blocked testing for absolute accuracy of global predictions was smaller for older than younger adults. However, individual differences in correlations of strategy effectiveness ratings and postdictions showed similar upgrades for both age groups. Older adults learn about imagery's superior effectiveness but do not accurately estimate the magnitude of its benefit, even after blocked testing.

Keywords: Metacognition; Strategy; Knowledge; Associative learning; Imagery.

INTRODUCTION

Optimal task performance often depends on whether individuals are able to identify and use the best strategy for a given task (e.g., Lemaire & Siegler,

This research was supported by Grant NIA R37 AG13148 awarded to C. Hertzog from the National Institute on Aging. Further information about the metacognition research conducted in the Hertzog lab can be found at <http://psychology.gatech.edu/CHertzog>.

Address correspondence to: Christopher Hertzog, School of Psychology, Georgia Institute of Technology, Atlanta, GA, 30332-0170, USA. E-mail: christopher.hertzog@psych.gatech.edu

1995; Schunn & Reder, 2001; Siegler & Stern, 1998). Knowledge updating (KU) reflects the ability to learn about the relative effectiveness of different strategies from task experience. Knowledge updating requires individuals to accurately monitor the differential effectiveness of the strategies during the task and then attribute differences in performance to the particular strategy used (Dunlosky & Hertzog, 2000; Hertzog, Price, & Dunlosky, in press-b).

Age-Related Differences and Equivalencies in Knowledge Updating

The limited number of studies that have examined age-related differences in KU have reached somewhat different conclusions. Knowledge updating has been measured by improved agreement of metacognitive judgments with the effects of various strategies (or conditions) after task experience. Brigham and Pressley (1988) asked younger and older adults to learn word-definition pairs using different strategies (e.g., the keyword method) and concluded that older adults were deficient in their ability to monitor and learn about differential strategy effectiveness.

Similarly, Bieman-Copland and Charness (1994) reported age-related differences in KU for item-level judgments of learning (JOLs) for the effectiveness of different types of cues for cued recall. Matvey, Dunlosky, Shaw, Parks, and Hertzog (2002) expanded on Bieman-Copland and Charness' (1994) work by including additional metacognitive judgments to assess KU. They measured global differentiated predictions and postdictions, which ask people to estimate the percentage of items studied under a given condition that would be (predictions) or had been (postdictions) remembered. Matvey et al. (2002) observed age differences in the changes across trials in absolute accuracy of global differentiated predictions and postdictions, with older adults showing less knowledge updating.

In contrast, Dunlosky and Hertzog (2000) did not find age differences in KU about the effects of imagery and repetition on paired-associate recall. They instructed younger and older adults to use both a normatively effective strategy (interactive imagery) and a normatively ineffective strategy (rote repetition) for learning new associations between normatively unrelated word pairs (e.g., TICK-SPOON), and measured changes in KU with metacognitive judgments. Their younger and older adults both exhibited KU as reflected by an increase in between-person correlations of global predictions with recall across trials, even though both groups badly underestimated their recall performance with interactive imagery.

Consider two possible reasons why age-related differences have not been consistently obtained. First, the resource demands for performance

monitoring required for KU may have been too great for both age groups, making it *appear* as if no age-related deficits exist in KU. The KU process is complex and demands resources (Bieman-Copland & Charness, 1994). Namely, one must encode paired associates using various strategies, and then at retrieval, recall which strategy had been used during encoding, associate test outcomes with strategies employed, and track how often each strategy produced successful recall. For both age groups, tracking recall outcomes while also performing on the recall test may exceed their attentional and working memory capacity (Bieman-Copland & Charness, 1994). If so, neither age group may be able to gain highly accurate information about the relative efficacy of the strategies, giving the *appearance* of age equivalence in KU. If the resource demands of the task were reduced to a level where older adults were taxed more than younger adults, then perhaps age-related differences in KU would emerge.

Second, the metacognitive judgments used to measure declarative knowledge about strategy effectiveness are influenced by multiple variables, not merely strategy effectiveness. In fact, work on judgments of learning suggests that they are far more influenced by observable stimulus characteristics (e.g., concreteness) than by processing strategies employed during encoding (Koriat, 1997; Lovelace, 1990). Hence, they may not be sufficiently sensitive to detect subtle age-related deficits. Similar issues may apply to the types of global (list-wise) predictions and postdictions used to study KU (Dunlosky & Hertzog, 2000; Matvey et al., 2002). If so, a more direct measure of declarative strategy knowledge may have better construct validity and could uncover age-related differences that are overshadowed by other sources of variance in metacognitive judgments. In the present study, we pursue both hypotheses regarding age differences in KU.

Hertzog et al. (in press-a) used blocked testing to lessen the demands on resources during the KU task, using only young adults as participants. As in Dunlosky and Hertzog (2000), individuals were instructed to use either rote repetition or interactive imagery to learn PA items. Concurrent item-level strategy reports were collected during encoding (Dunlosky & Hertzog, 1998), enabling a check on compliance with instructions, but also ensuring that the study strategy actually employed was known. Recall was tested using homogeneous blocks of items participants had studied with a particular strategy, while also explicitly informing participants at the start of the block about which strategy (imagery or repetition) they had originally reported using to study the items in that test block. With this procedure, participants no longer needed to remember which items were studied with which strategy at the time of item recall, and could more easily keep track of strategy success. As expected, younger adults' absolute accuracy of global judgments improved.

We hypothesized that if the limited age-related effects on KU in previous research resulted from resource limitations during test, then using blocked testing to inform individuals about their original encoding strategies should allow both age groups to gain more accurate information about the relative efficacy of imagery and repetition. However, older adults' greater processing resource limitations (relative to younger adults; e.g., Salthouse, 1991) might not allow them to benefit as much from blocked testing. This could occur, for example, if the PA learning task exhausted older adults' available resources, thus undermining any attempt to track performance gains according to strategy, even with blocked testing. Alternatively, older adults' associative learning deficit (e.g., Kausler, 1994; Naveh-Benjamin, 2000) could hinder forming associations between test outcomes and strategies used, resulting in less benefit from blocked testing, even when the strategy information was readily accessible.

Both age groups also completed a questionnaire measuring beliefs about the effectiveness of different strategies for learning new associations, including rote repetition and imagery (Hertzog & Dunlosky, 2004). Hertzog and Dunlosky (2006) showed that this measure predicted spontaneous strategy use during associative learning. Hence, administering the questionnaire before and after the task may provide a more sensitive measure of KU than metacognitive judgments.

To maximize the validity of the blocking manipulation it was necessary to block by the strategy used by participants rather than the instructed strategy because younger and older adults do not always comply with strategy instructions (e.g., Dunlosky & Hertzog, 1998). For instance, younger adults reported complying with imagery instructions on less than 80% of the trials. Hertzog et al. (in press-a) found that such rates of compliance actually affected estimates of KU, because individuals tended to shift to the more effective imagery strategy when instructed to use repetition, especially when studying a second list. Hence, as in Hertzog et al. (in press-a), blocking of items at test was done based on reported strategy use.

METHOD

Design and Participants

The design was a $2 \times 2 \times 2 \times 2$ mixed factorial, with Trial (first versus second study list), Strategy instructions (repetition versus imagery), Testing condition (random versus blocked), and Age group (younger versus older adults). Trial and Strategy instructions were manipulated within-subjects and Testing condition was a randomly assigned between-subjects factor.

Eighty-seven younger (48 males and 39 females) and 78 older adults (29 males and 49 females with $M = 15.28$, $SD = 3.0$ years of education)

participated in the experiment. Younger adults (M age = 19.2 years, SD = 1.3) were students at the Georgia Institute of Technology and received course credit for participating. Older adults (M age = 69.3 years, SD = 5.1) were normal, community-dwelling adults recruited from Atlanta, GA. They received a nominal fee for their participation. They had previously participated in a study of individual differences in skill acquisition, but had not been exposed to the type of paired-associate learning evaluated in this study. Random assignment to conditions resulted in 39 older and 43 younger adults in the random testing condition and 39 older and 44 younger adults in the blocked testing condition.

Materials

Demographic information was collected from participants using a brief background information questionnaire. The PEP questionnaire (see Hertzog & Dunlosky, 2004) was used prior to task exposure to measure participants' previously existing strategy knowledge as well as after the intervening experimental task to assess gains in strategy knowledge. The PEP lists various memory strategies (including imagery and repetition), along with their definitions and how they would be used. Participants rated the effectiveness of each strategy on a 10-point Likert-type scale provided below each strategy.

One hundred and twenty-four word pairs, consisting of relatively frequent, concrete nouns (e.g., TICK-SPOON) were used in this study. The word pairs were selected from the University of South Florida norms to have no prior association (Nelson, McEvoy, & Schreiber, 1998). Four of these word pairs were used during practice while the other 120 pairs were randomly divided into two lists of 60 word pairs. The experimental task was programmed in Visual Basic (Visual Studio, Version 6.0, Microsoft Corporation, 1998) programming language and run on PC desktop computers. All responses were entered and recorded on the computer keyboards.

Procedure

The experimental task consisted of two study-test lists, with 60 different word pairs presented for study on each list. Participants were asked to complete the background information form as well as the PEP I prior to receiving instructions about the PA experimental task. Participants were then told that they would be instructed to study half of the items with imagery and half with repetition and provide various metamemory judgments throughout the experiment. The task instructions included examples of how to use the imagery and repetition strategies as well as each type of metamemory judgment participants would be asked to make. After reading the instructions participants practiced studying and providing judgments for four word pairs, two of which they were instructed to study with imagery and the other two with repetition, so they would be familiar with the format.

Study

The presentation order for studying the items was randomized for each participant in each list. Younger adults were allotted 6 s and older adults 10 s to study each word pair to ensure older adults would not have floor recall performance for items studied with the normatively less effective rote repetition strategy.

For each list, the computer randomly assigned half of the PA to be studied using imagery and the other half with repetition. Participants were explicitly told to use one of the strategies for each item. A prompt appeared one second before each word pair appeared that instructed the participant to study the item with either 'Imagery' or 'Repetition'. The prompt remained on the screen during the allotted study time for the item. After study time elapsed for each item participants were asked to report which strategy, if any, they had *actually* used to study the item.

Test

After participants finished studying the 60 items in a given list they received instructions for the PA recall task and attempted to recall 40 of the 60 studied items. A subset of 40 items, rather than all 60, was used to ensure problems with strategy compliance would not prevent the formation of homogeneous blocks of recall testing for those in the blocked testing condition. The computer program was designed to select 20 items that had been studied with imagery and 20 studied with rote repetition. The same algorithm was used to select items in the random and blocked testing conditions. However, those in the random testing condition recalled the imagery and repetition items in a random order, whereas those in the blocked testing condition received four homogeneous blocks of 10 items each, and a prompt prior to the onset of each block regarding which strategy (imagery or repetition) they had reportedly used to study that set of items (e.g., 'You reported studying the following items using Imagery [Repetition].'). If participants were sufficiently noncompliant with instructed strategies to prevent the selection and formation of homogeneous blocks, then the prompt in the blocked testing condition informed participants that they had reported using a mixture of strategies to study the upcoming block of items.¹ For each item, the stimulus

¹The impact of noncompliance with instructed strategy use was examined by analyzing the data twice. First, the data were analyzed excluding the 10 older and three younger adult participants in the blocked testing condition that were sufficiently noncompliant to form homogeneous blocks of PA recall testing. These analyses were then compared to a second set of analyses run on the complete sample of participants. The loss of power in the reduced sample did yield minor changes in some significance tests, but substantive conclusions about KU were not altered. Because the overall pattern in the marginal means remained the same across both sets of analyses and because the inclusion of noncompliant participants would dilute rather than increase the effectiveness of blocked testing, we only report results from the complete sample.

was presented (e.g., ‘TICK–?’), and the participant was instructed to type the word that was originally paired with the stimulus (e.g., ‘spoon’). Participants had unlimited time to respond and omissions were not allowed. Responses were scored as correct if the first three letters matched the target response.

Metacognitive Judgments

A number of metacognitive judgments were collected in order to evaluate when KU occurs (for further discussion, see Dunlosky & Hertzog, 2000; Hertzog et al., in press-a, in press-b). Judgments were requested prior to study (global prestudy predictions), during study (judgments of learning), after study but before the recall test (global poststudy predictions), and after the test (global postdictions).

Immediately before beginning a study-test trial, participants were told they would be asked to study 30 items with imagery and 30 with rote repetition and were asked to predict what percentage of items studied with each strategy they would be able to recall. Specifically, participants were asked to make a prediction for each strategy by typing ‘any number between 0 and 100 that corresponds to the percentage of pairs that you will study using Interactive Imagery [or Rote Repetition] that you will correctly recall’. Similar predictions were collected after participants finished studying the 60 items within each list, but before the recall test, to allow assessment of changes in predictions as a result of encoding experience. Because these predictions do not add materially to our treatment of blocking effects on knowledge updating, we shall not report data on them in this paper.

During the study phase, participants were asked to provide JOLs immediately after the offset of the presentation of each word pair. Participants were shown only the stimulus of an item (e.g., if ‘tick–spoon’ had been presented for study, the prompt would include ‘tick–?’) and the query, ‘How confident are you that in about ten minutes from now you will be able to recall the second word of the pair when prompted with the first?’ Participants were asked to type any number between 0 and 100, with 0 indicating they were sure they would not be able to recall the second word when prompted with the first, and 100 indicating they were 100% certain they would be able to recall the second word 10 min from the time of their judgment.

After completing paired associate recall for all 40 items in a list, participants were immediately asked to provide three postdictions. The first was a global postdiction, without reference to the type of strategy used. This type of postdiction has been shown to correlate highly with performance on recall tests (e.g., Devolder, Brigham, & Pressley, 1990; Hertzog et al., in press-b; Hertzog, Saylor, Dixon, & Fleece, 1994). We refer to this postdiction as an undifferentiated postdiction (with respect to strategy). We also collected two separate global-differentiated postdictions, one for items studied with

interactive imagery and another for items studied with rote repetition. For each postdiction, individuals were asked to type a number between 0 and 100 to indicate what percentage of items had been recalled. For the global-differentiated postdictions, participants estimated recall of items studied with each type of strategy.

Strategy Effectiveness Questionnaire

Upon completing the two study-test trials and the postdictions for the second list, participants were again asked to complete the PEP to assess updating of declarative knowledge about strategy effectiveness.

RESULTS

In all the results that follow, dependent variables were analyzed in a $2 \times 2 \times 2 \times 2$ (Age by Test by Strategy by Trial) general linear model (GLM), with repeated measures on Strategy and Trial.

Compliance with Strategy Instructions

Participants complied with 75% of strategy instructions for studying items in both the first and second list, but with greater imagery (marginal $M = 84.3\%$, $SE = 1.2$) than repetition (marginal $M = 65.3\%$, $SE = 1.8$) compliance, $F(1, 161) = 97.40$, $p < .001$, partial $\eta^2 = .377$, and higher compliance by younger (marginal $M = 77.6\%$, $SE = 1.7$) than older adults (marginal $M = 72.6\%$, $SE = 1.8$), $F(1,161) = 5.21$, $p < .05$, partial $\eta^2 = .031$. Participants' compliance rates changed across lists, producing a significant Trial \times Strategy \times Age interaction, $F(1, 161) = 5.55$, $p < .05$, partial $\eta^2 = .033$. Younger adults became more compliant with imagery (List 1 marginal $M = 82.6\%$, $SE = 1.6$ versus List 2 $M = 84.4\%$, $SE = 1.8$), but less compliant with repetition instructions across lists (List 1 marginal $M = 73.2\%$, $SE = 2.6$ versus List 2 $M = 70.1\%$, $SE = 2.9$). This replicates findings from Hertzog et al. (in press-a). Older adults showed the opposite pattern. They became more compliant with repetition (List 1 marginal $M = 58.3\%$, $SE = 2.7$ versus List 2 $M = 59.6\%$, $SE = 3.1$) and less compliant with imagery instructions (List 1 marginal $M = 86.5\%$, $SE = 1.7$ versus List 2 $M = 83.8\%$, $SE = 1.9$) across lists. Thus, whereas younger adults showed a shift away from repetition and toward greater imagery compliance, older adults showed a slight shift away from imagery and toward repetition use across lists. The reason for older adults' shift away from imagery use is unclear and could be due to problems forming images, which could have resulted in slightly greater reliance on the repetition strategy, or their inability to use any strategy on some of the PA items. Note, however, that the changes in compliance were on average rather small for older adults (less than a 3% change for either strategy). No other effects were significant.

Because compliance rates were less than perfect, data were analyzed solely as a function of reported strategy use (see Hertzog et al., in press-a for comparisons of reported versus instructed strategy use, and further discussion of why reported strategy use is the proper basis of analysis).

Recall Performance

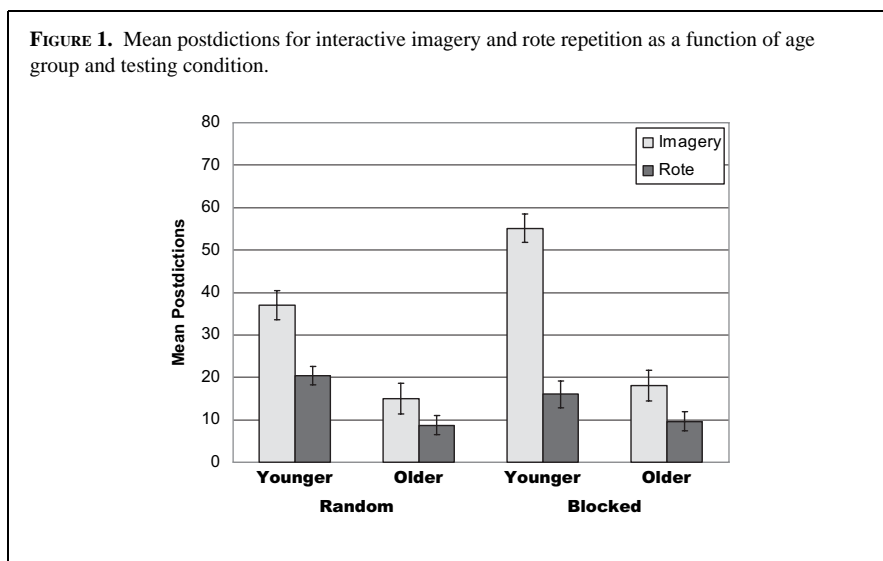
The differential presentation rates prevented older adults from having floor recall performance for repetition items. Yet, older adults (M imagery = 31%, $SE = 2.4$ versus M repetition = 14%, $SE = 1.9$) had significantly lower recall levels with both types of strategies than did younger adults (M imagery = 64%, $SE = 2.3$ versus M repetition $SE = 25%$, $SE = 1.8$), $F(1, 157) = 50.74$, $p < .001$, partial $\eta^2 = .244$. Critically for any evaluation of KU, recall for imagery items was reliably greater than repetition recall in both age groups, $F(1, 157) = 359.71$, $p < .001$, partial $\eta^2 = .696$. The clear superiority of the imagery strategy makes it possible to study KU by examining whether participants' metamemory judgments showed knowledge of differential strategy effectiveness (see Dunlosky & Hertzog, 2000, for further discussion). More important, the nonsignificant Strategy \times Testing condition \times Age group interaction, $F(1, 157) = 2.11$, $p > .10$, partial $\eta^2 = .013$, indicated that blocked testing did not affect age differences in strategy effectiveness. Thus, any blocking effect on absolute accuracy could not be attributed to effects of the manipulation on recall.

Postdictions and Postdiction Accuracy

The blocked testing manipulation was expected to have the greatest influence on participants' global postdictions because they were collected immediately after exposure to either random or blocked testing for both List 1 and List 2. To the extent that blocking results in better monitoring of relative strategy effectiveness, the postdictions in the blocked testing condition should be more accurate than postdictions in the random condition. For this reason the postdiction data are reported first.

Global Differentiated Postdictions

The effects of blocking on the postdictions were similar for both lists (there were no interactions involving Trial), so we report blocking effects aggregating over lists. The critical effect was a reliable Strategy \times Testing condition \times Age group interaction, $F(1, 161) = 11.61$, $p < .001$, partial $\eta^2 = .067$. As can be seen in Figure 1, blocked testing had no impact, relative to random testing, on older adults' postdictions. Younger adults, in contrast, showed a significantly greater separation between their imagery and repetition judgments after blocked than random testing, (M difference = 39.16, $SE = 2.95$ versus M difference = 16.64, $SE = 2.99$, respectively). When each age group's data were analyzed separately, younger adults' postdictions produced



a reliable Strategy \times Testing condition interaction, $F(1, 85) = 19.76, p < .001$, partial $\eta^2 = .189$, whereas older adults' postdictions did not, $F(1, 76) = 0.63, p > .10$, partial $\eta^2 = .008$. The similarity across lists indicated that a single study-test trial was sufficient for younger adults to learn about differential strategy effectiveness. In sum, blocked testing facilitated younger adults' learning of imagery's superiority to repetition, but the additional support afforded by blocked testing did not seem to help older adults.

Absolute Accuracy

Of interest then was whether blocked testing would yield greater absolute accuracy of younger adults' postdictions, calculated as the difference between participants' postdicted and actual recall.² Table 1 reports the absolute accuracy data. Postdiction accuracy actually decreased slightly across trials, $F(1, 157) = 4.26, p < .05$, partial $\eta^2 = .026$, (List 1 marginal $M = -9.81, SE = 1.04$ versus List 2 marginal $M = -11.76, SE = 1.08$) and was reliably worse for imagery (marginal $M = -16.08, SE = 1.28$) than repetition

²The analyses focused on the absolute accuracy of participants' global predictions, JOLs, and postdictions because it was more important to know in an absolute sense whether individuals were able to track the differential effectiveness of imagery and repetition at a global level across lists (as measured by the absolute accuracy of the global metamemory judgments). Experimental studies of metacognitive monitoring often assess the relative accuracy of participants' item-level JOLs. We evaluated whether participants' JOLs were able to differentiate items that would and would not be recalled as a function of strategy use (assessed with relative accuracy, as measured by gamma correlations), but do not report these data here. Contact the first author for a more complete set of results if interested.

TABLE 1. Absolute accuracy of global postdictions

Age group	List 1				List 2			
	Imagery		Repetition		Imagery		Repetition	
	<i>M</i>	<i>(SE)</i>	<i>M</i>	<i>(SE)</i>	<i>M</i>	<i>(SE)</i>	<i>M</i>	<i>(SE)</i>
<i>Random</i>								
Younger	-21.82	(2.8)	-3.93	(2.3)	-25.12	(2.9)	-11.13	(2.6)
Older	-17.82	(2.9)	-5.97	(2.4)	-16.12	(3.0)	-8.69	(2.8)
<i>Blocked</i>								
Younger	-12.23	(2.7)	-5.03	(2.3)	-10.81	(2.8)	-5.98	(2.6)
Older	-12.15	(2.9)	0.49	(2.4)	-12.57	(3.0)	-3.62	(2.8)

Note: Entries are means (and standard errors) of individuals' difference scores between global differentiated postdictions and recall. Imagery, items for which participants reported using imagery to study the items; Repetition, items for which participants reported using rote repetition to study the items.

items (marginal $M = -5.48$, $SE = 1.09$), $F(1, 157) = 55.00$, $p < .001$, partial $\eta^2 = .259$. Accuracy for imagery items was the same across lists (List 1 marginal $M = -16.00$, $SE = 1.42$ versus List 2 $M = -16.16$, $SE = 1.46$), and remained worse than accuracy for repetition items, which decreased across lists (List 1 $M = -3.61$, $SE = 1.18$ versus List 2 $M = -7.36$, $SE = 1.34$), thereby yielding a reliable Trial \times Strategy interaction, $F(1, 157) = 4.12$, $p < .05$, partial $\eta^2 = .026$.

Most important for the blocking hypothesis was whether underestimation of imagery strategy effectiveness was influenced by blocked testing. The Strategy \times Testing condition \times Age group interaction just missed conventional significance, $F(1, 157) = 3.76$, $p = .054$, partial $\eta^2 = .023$, reflecting the trend for younger adults to be more accurate than older adults, but participants in both age groups to have better absolute accuracy after blocked than random testing. None of the aggregate higher-order interactions was reliable. However, analyses run separately for each age group revealed the Strategy \times Testing interaction was reliable for younger adults, $F(1, 83) = 4.43$, $p < .05$, partial $\eta^2 = .051$, but not for older adults, $F < 1$. Older adults did show an effect of blocked testing, $F(1, 74) = 4.52$, $p < .05$, partial $\eta^2 = .058$, with better absolute accuracy for blocked testing (marginal $M = -7.0$, $SE = 1.7$) than random testing ($M = -12.2$, $SE = 1.7$), but this effect did not affect differential accuracy for the two strategies.

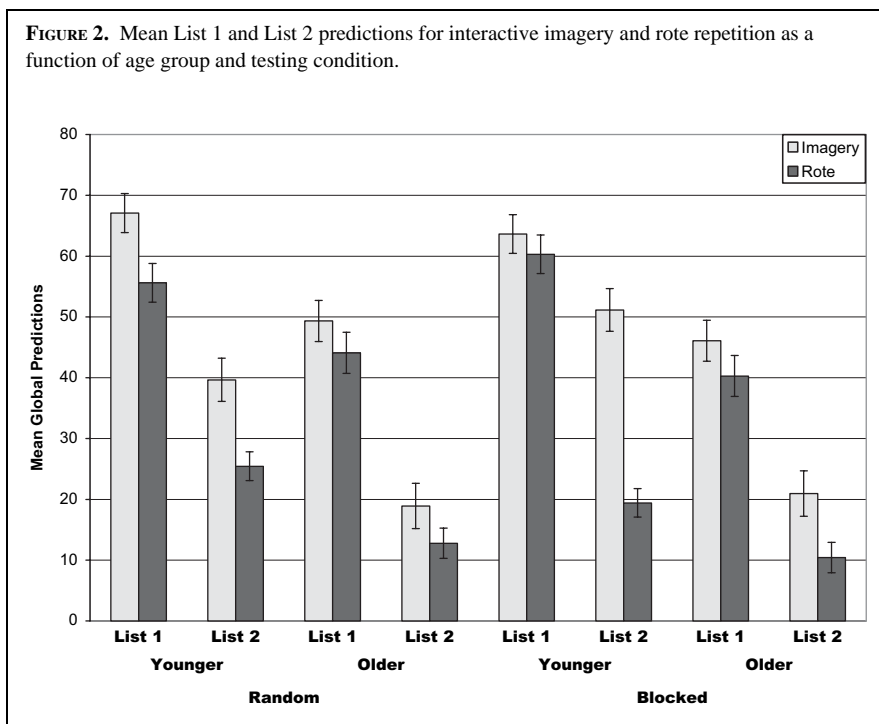
Thus in the aggregate, postdiction absolute accuracy was greater after blocked than random testing, despite testing condition having no impact on the differential accuracy of repetition and imagery postdictions. There was one other indication that blocked testing affected younger adults more than older adults. Younger adults in the random condition actually showed worse

postdiction accuracy on the second list, but this was not the case for the younger adults in the blocked testing condition, yielding a Trial \times Test interaction $F(1, 83) = 4.07, p < .05$, partial $\eta^2 = .047$. Older adults given blocked testing and participants in the random testing condition did not show any tendency for accuracy to improve over lists to make up for their worse postdiction accuracy at List 1, $F < 1$.

Predictions and Prediction Accuracy

Global Differentiated Predictions

Blocking effects occurred on List 1 postdictions. Hence, an important question is whether these positive effects would carry over to List 2 predictions (see Figure 2). The GLM yielded a reliable four-way interaction of Trial \times Strategy \times Age group \times Testing condition, $F(1, 161) = 9.97, p < .01$, partial $\eta^2 = .058$. Individuals in both testing conditions and age groups lowered their predictions for both types of strategies in List 2, relative to their predictions in List 1. However, individuals in the random testing condition lowered their predictions for both types of strategies to a similar extent such that the mean difference between how much imagery and repetition predictions were downgraded across lists was only 0.9 for older adults (M imagery reduction = 30.41, $SE = 3.34$ versus M repetition reduction = 31.31, $SE = 2.93$)



and 2.73 for younger adults (M imagery reduction = 27.44 versus M repetition reduction = 30.17). In contrast, younger adults exposed to blocked testing differentially downgraded their predictions, reducing repetition predictions to a much greater extent than imagery predictions (M imagery reduction = 12.50, $SE = 3.35$ versus M repetition reduction = 40.89, $SE = 2.76$), thereby yielding an overall difference of 28.39. This same pattern was not observed for older adults' List 2 predictions in the blocked condition, who reduced imagery and repetition predictions to a similar extent (M imagery reduction = 25.13, $SE = 3.56$ versus M repetition reduction = 29.87, $SE = 2.93$), resulting in an overall difference of 4.74.

When analyzed separately in each age group, these differences in downgrading patterns across the two testing conditions were reliable for younger adults, yielding a reliable Trial \times Strategy \times Testing condition interaction, $F(1, 85) = 22.38$, $p < .001$, partial $\eta^2 = .208$. This interaction was not present for older adults, $F < 1$. Thus, although older adults' global predictions did benefit slightly from blocked testing, the adjustment of their ratings did not create the greater separation for imagery and rote repetition strategies found for the younger adults exposed to blocked testing.

Absolute Accuracy

Table 2 reports the absolute accuracy of participants' global predictions as a function of age group and testing condition. Individuals in the blocked testing condition showed greater separation in the downgrading of their List 2 predictions for imagery and repetition than those in the random testing condition, which yielded a reliable interaction of Trial, Strategy, and Testing condition, $F(1, 157) = 3.92$, $p = .05$, partial $\eta^2 = .024$. Younger adults showed greater separation in their downgraded predictions than older adults, producing a reliable interaction of Trial, Strategy, and Age group, $F(1, 157) = 5.85$, $p < .05$, partial $\eta^2 = .036$. Examination of the marginal means suggests that it was the younger adults' data driving the differences between testing conditions. Older adults showed similar patterns in the absolute accuracy of their predictions across both testing conditions, whereas the younger adults showed greater divergence in their imagery and repetition ratings after exposure to blocked testing. However, the four-way interaction of Trial, Strategy, Testing condition, and Age group was not reliable, $F < 1$. In sum, the global predictions suggested that younger adults benefited more than older adults from blocked testing, in terms of absolute accuracy of List 2, relative to List 1, predictions differentiating between the imagery and rote strategies.

JOLs

Mean JOLs across age groups and testing conditions are displayed in Figure 3. Even at the first list, individuals from both age groups gave higher JOLs to items studied with imagery than to items studied with rote. These

TABLE 2. Absolute accuracy of global predictions and judgments of learning

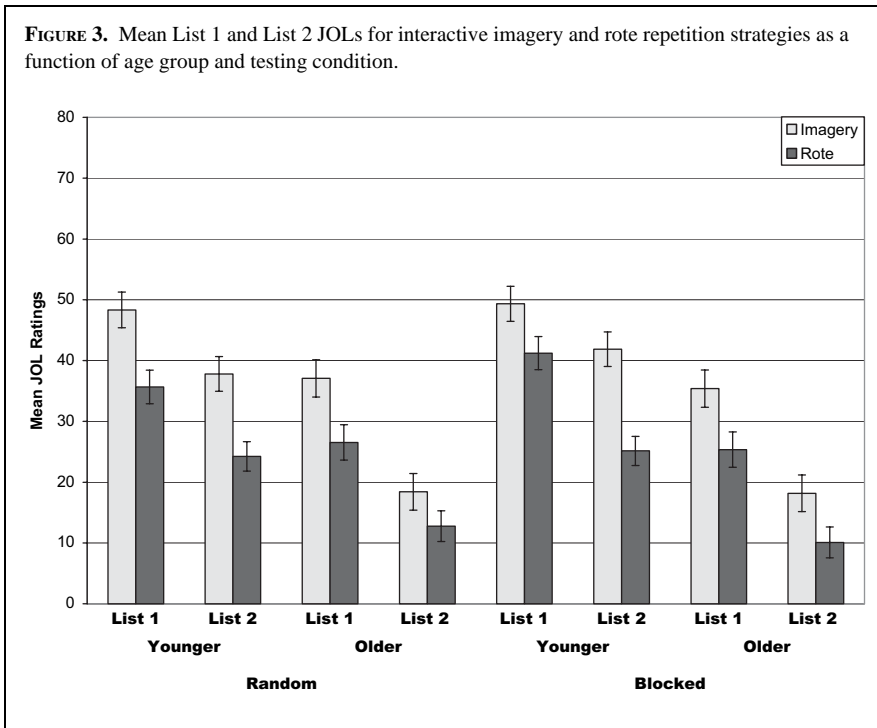
Age group	List 1				List 2			
	Imagery		Repetition		Imagery		Repetition	
	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)
Random								
<i>Younger</i>								
Global predictions	6.13	(3.8)	29.60	(4.0)	-20.88	(3.2)	-5.25	(3.0)
JOLs	-13.05	(4.0)	10.12	(3.5)	-23.63	(3.6)	-7.01	(3.2)
<i>Older</i>								
Global predictions	15.76	(4.0)	28.24	(4.2)	-11.65	(3.4)	-3.82	(3.2)
JOLs	3.50	(4.2)	10.86	(3.7)	-12.40	(3.8)	-3.91	(3.4)
Blocked								
<i>Younger</i>								
Global predictions	-3.79	(3.8)	32.28	(3.9)	-17.02	(3.2)	-1.10	(3.0)
JOLs	-17.71	(4.0)	18.43	(3.5)	-25.15	(3.5)	4.20	(3.2)
<i>Older</i>								
Global predictions	15.35	(4.0)	29.33	(4.2)	-9.39	(3.4)	-1.20	(3.2)
JOLs	4.77	(4.2)	14.67	(3.7)	-12.21	(3.8)	-1.28	(3.4)

Note: Entries are means (and standard errors) of individuals' difference scores between predictions (either judgments of learning or global differentiated predictions) and recall. Imagery, items for which participants reported using imagery to study the items; Repetition, items for which participants reported using rote repetition to study the items.

differences appeared to be less influenced by task experience than the meta-cognitive judgments already reported. Nevertheless, participants' JOLs revealed reliable Trial \times Strategy \times Testing condition, $F(1, 157) = 8.05$, $p < .01$, partial $\eta^2 = .049$, and Trial \times Strategy \times Age interactions, $F(1, 157) = 19.03$, $p < .001$, partial $\eta^2 = .108$. For both testing conditions, younger adults tended to increase the difference in mean JOLs slightly across the two lists, whereas older adults did not. More critically, imagery and repetition JOLs showed slightly greater separation in List 2 after blocked testing for younger adults (marginal mean differences in JOLs for the two strategies increased in List 2 by 8% confidence in the blocked condition, 2% in the random condition). Conversely, older adults reduced the differences by -2% for the blocked condition, and -5% for the random condition on List 2, relative to List 1.

JOL Accuracy

The absolute accuracy of JOLs was examined by calculating the difference between each individual's mean JOLs and their mean level of actual recall performance (see Table 2). Reliable main effects were observed for Trial, $F(1, 157) = 96.28$, $p < .001$, partial $\eta^2 = .380$, and Strategy, $F(1, 157) = 161.79$, $p < .001$, partial $\eta^2 = .508$. In addition, significant interactions



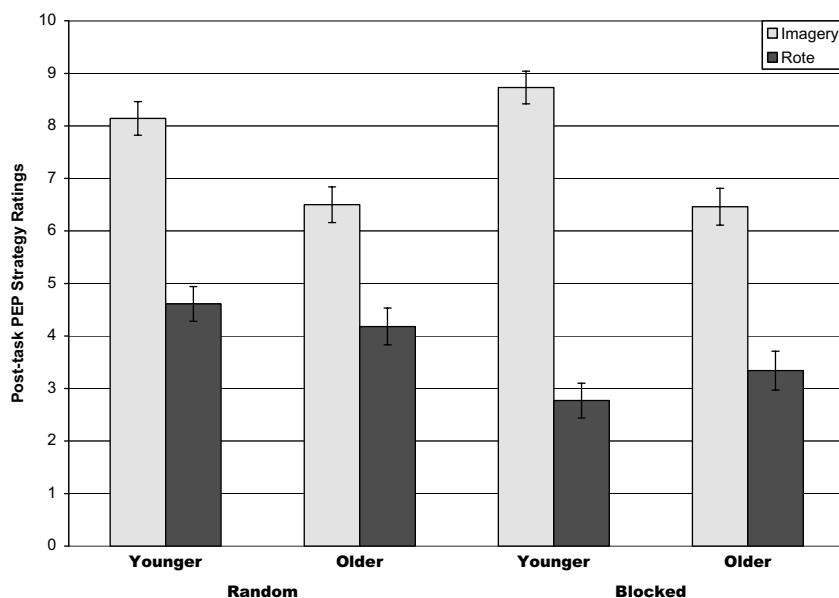
were found between Strategy \times Testing condition, $F(1, 157) = 7.55, p < .01$, partial $\eta^2 = .046$, and Strategy \times Age, $F(1, 157) = 37.77, p < .001$, partial $\eta^2 = .194$. A three-way interaction between Trial \times Strategy \times Age also was reliable, $F(1, 157) = 4.68, p < .05$, partial $\eta^2 = .029$. Marginal means indicated that the absolute accuracy of participants' JOLs decreased across lists ($M = 3.95, SE = 1.74$ in List 1 versus $M = -10.17, SE = 1.54$ in List 2) and was worse for imagery ($M = -11.99, SE = 1.78$) than repetition items ($M = 5.76, SE = 1.47$). Individuals' JOLs in both testing conditions underestimated their imagery and overestimated their repetition recall performance. Surprisingly those in the random testing condition had better absolute accuracy than those in the blocked testing condition. Similarly, both younger and older adults' JOLs underestimated imagery but overestimated repetition recall performance. However, older adults' JOLs more closely aligned with actual recall levels across both lists, and when collapsed across lists as a result of their having lower recall. Thus, individuals in both age groups acted to lower JOLs drastically for both types of items, producing better absolute accuracy, but without necessarily reflecting the underlying superiority of the imagery strategy. The overall impression from the data is that blocking had a weak effect on JOLs and did not evidence any signs of differential KU for List 2.

Strategy Effectiveness Questionnaire

Our analyses of the strategy questionnaire (i.e., PEP) data focused only on the two instructed strategies (i.e., imagery and repetition), before and after task experience. PEP ratings could range from 0 (completely ineffective strategy) to 10 (completely effective strategy). Before task experience, imagery and rote repetition were rated as equally effective strategies by both young adults (imagery $M = 6.23$, $SE = 0.30$; repetition $M = 6.47$, $SE = 0.23$) and older adults (imagery $M = 6.59$, $SE = 0.28$, repetition $M = 6.66$, $SE = 0.21$). If anything, the slight trend for the pre-experimental questionnaire ratings was for better ratings for rote repetition over imagery in both age groups.

This pattern changed dramatically after task experience (see Figure 4). The GLM revealed that both younger and older adults rated imagery higher than repetition (Younger M imagery rating = 7.51, $SE = 0.21$, and M repetition rating = 5.18, $SE = 0.18$ versus Older M imagery rating = 6.36, $SE = 0.23$ versus M repetition rating = 5.12, $SE = 0.19$), but the greater separation in younger adults' ratings resulted in a reliable Strategy \times Age group interaction, $F(1, 156) = 6.47$, $p < .05$, partial $\eta^2 = .040$. On average, participants' imagery ratings increased (List 1 $M = 6.41$, $SE = 0.21$ versus List 2 $M = 7.46$, $SE = 0.17$), whereas repetition ratings decreased across lists (List 1 $M = 6.57$,

FIGURE 4. Mean post-task PEP II strategy effectiveness ratings for interactive imagery and rote repetition strategies as a function of age group and testing condition.



$SE = 0.16$ versus List 2 $M=3.73$, $SE = 0.17$), which produced a reliable Trial \times Strategy interaction, $F(1, 156) = 180.58$, $p < .001$, partial $\eta^2 = .537$. These interactions were qualified by the significant Trial \times Strategy \times Age group, $F(1, 156) = 10.46$, $p < .001$, partial $\eta^2 = .063$, interaction which indicated greater separation between the two strategies after task experience for younger (M separation = 4.81, $SE = 0.24$) than older adults (M separation = 2.96, $SE = 0.24$). The reliable three-way interaction of Trial \times Strategy \times Testing condition, $F(1, 156) = 6.91$, $p < .01$, partial $\eta^2 = 0.042$, further demonstrated that persons in the blocked testing condition showed greater separation between their imagery and repetition ratings (M separation = 3.64, $SE = 0.25$) than those in the random testing condition (M separation = 2.69, $SE = 0.25$) after task experience. Finally, a reliable Trial \times Strategy \times Testing condition \times Age interaction, $F(1, 156)=4.27$, $p < .05$, partial $\eta^2 = .027$, indicated that the effect of blocking on strategy effectiveness ratings was larger for younger adults than for older adults.

Figure 4 clearly shows that both age groups show evidence of knowledge updating after a starting point of equivalent imagery and rote ratings, but the effect of blocking is clearly larger for younger adults. Thus, both age groups apparently learned through task experience that imagery was more effective for learning than repetition. Both age groups benefited from blocked testing, as indicated by younger and older adults both showing greater separation in their imagery and repetition ratings after exposure to blocked than random testing. However, this effect was substantially larger for the younger adults.

Individual Differences in Knowledge Updating

Correlations of Metacognitive Judgments with Recall Performance

Although the absolute accuracy of global predictions and postdictions did not suggest much knowledge updating in older adults, correlational measures could tell a different story. Dunlosky and Hertzog (2000) and Hertzog et al. (in press-b) found evidence of changes in correlations of predictions, postdictions, and the effectiveness ratings with recall even under conditions that produced little evidence of updating in absolute accuracy of global judgments and JOLs.

These data also manifested increases in correlations of global predictions with recall from List 1 to List 2. The correlation of the imagery global prediction with recall increased from .38 for List 1 to .64 for List 2. The correlation of the rote global prediction with recall increased from $-.01$ for List 1 to .49 for List 2. As in earlier studies, this increase was associated with accurate List 1 postdictions, which correlated .70 for imagery recall and .52 with rote recall. This pattern did not differ appreciably between the two age groups. The increase in correlations is consistent with the view that accurate

performance monitoring, as reflected in strong correlations of postdictions with recall, drives the increase in prediction accuracy (Finn & Metcalfe, 2007; Hertzog et al., in press-b).

Here we focus on evaluating the evidence for KU as measured by strategy effectiveness ratings before and after task experience. Hertzog et al. (in press-b) showed that a measure of knowledge updating, the difference in strategy effectiveness ratings for imagery and rote strategies, was reliably predicted by differences in postdictions for those two strategies in regression models in two samples of university students.

We computed the same variable in this study, the difference in effectiveness ratings for the PEP questionnaire before and after task experience. We also computed the difference between reported imagery and rote strategy recall and the difference between global postdictions for the two strategies. Table 3 reports the correlations between these difference score measures, broken down by test condition and age group. The pattern in the correlations was clear. In the blocked condition, both younger and older adults showed large increases in correlations of differences in strategy effectiveness ratings with the differences in recall for the two strategies. In the random condition, the correlations increased slightly, but did not approach the magnitudes achieved under blocked testing. In contrast to the absolute accuracy measures reported earlier, there was no evidence of age differences in this pattern of correlational upgrade. Thus, individual differences in the magnitudes of rated differences between the strategies were closely aligned with actual recall differences only in the blocked condition.

Regression Model for Updated Strategy Knowledge

We extended the correlational results by running a generalized linear regression model in SPSS with the difference in strategy effectiveness ratings as the dependent variable. We also used the average (over the two lists) difference in recall (imagery – rote) as an independent variable, along with the pre-experimental difference in strategy effectiveness ratings. Using the

TABLE 3. Correlations of the imagery – rote recall difference score with imagery – rote difference scores for strategy effectiveness ratings and postdictions as a function of age and testing condition

Age, Condition	PEP1 (imagery – rote)	Postdictions (imagery – rote)	PEP2 (imagery – rote)
Young, Random	-.08	.24	.10
Young, Blocked	.10	.76**	.69**
Old, Random	.16	.34*	.32*
Old, Blocked	.31	.80**	.63**

Abbreviations: PEP, Personal Encoding Preference Questionnaire.

* $p < .05$; ** $p < .001$.

TABLE 4. Generalized Linear Model regression coefficients predicting strategy effectiveness ratings after practice (PEP2) as a function of testing condition (blocked versus random), pre-experimental imagery-rote strategy effectiveness ratings, imagery-rote recall differences, and age

Source	Estimate (SE)	Wald LR χ^2 test
Intercept	2.12 (0.59)	12.69**
Test	0.60 (0.62)	0.93
PEP1 imagery – rote	0.25 (0.05)	23.07**
Recall imagery – rote	8.72 (1.26)	47.56**
Test \times recall	- 5.96 (1.71)	12.20**
Age	- 0.62 (0.43)	2.08

** $p < .001$.

latter variable as a predictor allowed us to make inferences about whether the other independent variables accounted for changes in strategy knowledge as a function of task experience. Table 4 reports the parameter estimates and significance tests for the model. As would be expected from the correlations just reported, the Test \times Recall difference interaction was significant, as were the linear effects for recall and the pre-experimental difference in strategy effectiveness ratings. Age was not significant. When all other age interactions were added to the model, none were reliable ($p > .25$).

We also ran a model that used the difference in postdictions for the imagery and rote strategies as an additional predictor variable. This model allowed us to evaluate whether subjective strategy efficacy, as influenced by performance monitoring, carried much of the effect of recall differences on the KU effect. The difference in postdictions was a potent predictor of post-experimental strategy knowledge, controlling on pre-experimental knowledge and actual recall differences (see Hertzog et al., in press-b, for similar results). This outcome supported the hypothesis that postdictions mediate much of the effect of actual recall differences between the strategies on KU. Again, when interaction effects involving age were added to the model, none were significant.

Working Memory as a Predictor of KU Effects

Correlational data were also relevant to the issue of whether cognitive resource limitations helped to produce older adults' poorer knowledge updating, even with blocked testing. We did not collect measures of cognitive resources such as working memory as part of this study. However, because our older participants had been recruited from a study on individual differences in skill acquisition, measures of working memory (WM) were available for 76 of our older participants. We used a z -score composite of these measures – Salthouse and Babcock's (1991) Computation Span and

TABLE 5. Correlations of working memory, perceptual speed, and vocabulary with imagery – rote recall differences, imagery – rote postdiction differences, and imagery – rote effectiveness ratings after task experience in the total older sample and by testing condition

Measure	Sample	Working memory	Perceptual speed	Vocabulary
I-R PEP1	Total	-.08	.06	.05
	Random	-.13	.18	.05
	Blocked	-.04	-.05	.06
I-R Recall	Total	.09	.26*	.27*
	Random	-.09	.15	.23
	Blocked	.26	.37*	.30
I-R Post	Total	.08	.07	.24*
	Random	-.25	-.33*	.17
	Blocked	.29	.32*	.27
I-R PEP2	Total	.05	.03	.34**
	Random	-.22	-.15	.29
	Blocked	.33	.21	.37*

Abbreviations: I-R, imagery – rote; PEP, Personal Encoding Preferences Questionnaire; Post, postdiction.

* $p < .05$; ** $p < .001$.

Reading Span – to evaluate whether older adults with lower levels of WM were less likely to manifest knowledge updating. Given their central relevance to knowledge updating, we used global postdictions and the strategy knowledge questionnaires as dependent variables.

To our surprise, the composite WM variable did not correlate reliably with the absolute accuracy of imagery and rote postdictions, the difference between imagery and rote postdictions, or the difference in strategy effectiveness ratings (see Table 5). A composite measure of perceptual speed, using Salthouse and Babcock's (1991) Pattern Comparison and Letter Comparison tests, also was not correlated with these variables. The Advanced Vocabulary test (Ekstrom, French, Harman, & Dermen, 1976) did correlate with the differences, ruling out the hypothesis that limited reliability of the difference score measures suppressed relationships to the WM variable.

DISCUSSION

In order for knowledge updating to occur in list-learning tasks, individuals must monitor which strategies are used to study particular list items and then attribute later recall outcomes to those encoding strategies. This study suggests that standard testing formats with long item lists create difficulties in individuals monitoring performance outcomes and associating those outcomes with the strategies originally used at encoding (see also Hertzog et al., in press-a). Blocked testing removes the need to associate strategy outcomes

with past study behavior, and enhances the sensitivity of performance monitoring (as measured by postdictions for each strategy) to recall differences between the two strategies. In turn, it also creates greater strategy differentiation of performance predictions for the second list and post-task effectiveness ratings. Most important for the present paper, older adults showed reduced benefits of blocked testing on metacognitive judgments compared to younger adults.

The major source of KU effects appears to be accurate performance monitoring (Hertzog et al., in press-b). Individuals in cued recall tasks are accurate in monitoring whether provided answers are correct (e.g., Higham, 2002), and this accurate performance monitoring carries over to relatively accurate postdictions. At the level of individual differences, accurate postdictions are the prime predictor of changes in strategy effectiveness ratings. However, the fact that postdiction magnitudes do not fully reflect the benefits of the imagery strategy, especially when items are tested in a random order, shows that individuals can accurately monitor overall levels of recall performance while experiencing some distortion in the way in which they attribute recall success to the two strategies (see Hertzog et al., in press-b, for further discussion). Differentiated postdictions require inferential mechanisms to translate accurate performance monitoring into an accurate subjective sense of the level of performance achieved with each strategy, which in turn determines knowledge updating.

The differential blocking effects on the absolute accuracy of global predictions and postdictions is consistent with the hypothesis that older adults' learning about strategy effectiveness is constrained by resource demands, as hypothesized by Bieman-Copland and Charness (1994). Blocked testing had a major impact on younger adults' postdictions, which carried over to their List 2 predictions. Thus, it appears that Dunlosky and Hertzog (2000) failed to find age differences in knowledge updating in their imagery-rote instruction experiment because the requirements of the recall test suppressed monitoring performance in a way that would permit accurate estimates of recall performance for each of the two strategies. Our recent work on this problem indicates that informing individuals that the goal of the experiment is to learn about differential strategy effectiveness, and instructing them to count successes and failures after using the two strategies, does not affect this phenomenon (Hertzog et al., in press-a). Any intent to monitor precisely the outcomes of all item recall attempts, while correlating that with the strategy previously used for those items, is probably overridden by the difficulty of maintaining accurate representations of rates of success while simultaneously searching for targets associated with new test cues. Blocked testing ameliorates this problem, differentially for younger adults.

However, one should not discount the fact that older adults do rate imagery as more effective than repetition after task experience. They also

show an increase in the correlation of strategy effectiveness rating differences with recall differences in the blocked testing condition that is similar to younger adults. This outcome suggests that older adults do indeed learn in a generic sense that imagery is a more effective strategy after task experience; more so in the blocked testing condition than in the random testing condition. However, older adults do not show the same degree of change in effectiveness ratings.

Why the age differences in absolute accuracy of postdictions and predictions after blocked testing? The resource-reduction hypothesis states that older adults are so engaged in the primary task – retrieval search after cueing – that they do not use the blocked testing to simultaneously monitor the benefits of different strategies for recall across different test blocks. The correlational upgrade data seem to make a general resource-reduction account of age differences in knowledge updating less plausible. Older adults benefit from blocked testing because their effectiveness ratings are more highly correlated with actual recall differences after blocked testing. Moreover, measures of WM do not correlate with KU effects in our older sample. This latter finding suggests that it is not a structural constraint on available WM resources that accounts for the differential effects on absolute accuracy after blocked testing. However, it could still be the case that, given the primary task demand imposed by the retrieval search during the recall test, older adults are less likely to use available resources strategically to effectively monitor performance outcomes. In effect, the problem could be one of using the available resources when under the dynamic load in the task context itself. Alternatively, the resources that are available to older adults may be sufficient to gain a general subjective impression of imagery superiority but insufficient to support a monitoring mechanism that would enable accurate quantitative estimates of strategy effectiveness. Finally, working memory resources may not be as critical to KU as executive functioning resources related to divided attention or to source monitoring while under attentional load.

One way a new experiment could address this issue would be to introduce a divided attention condition at retrieval (e.g., Craik, Naveh-Benjamin, Ishaik, & Anderson, 2000). If divided attention suppresses the blocked testing effect in younger adults as well as older adults, eliminating the relevant improvements in the absolute accuracy of metacognitive judgments and the questionnaire ratings, it would implicate an on-line process of monitoring performance outcomes that demands attentional resources. An interesting question is whether divided attention would preserve the correlational shifts seen during blocked testing while suppressing the improvements in absolute accuracy for global predictions and postdictions. Such an outcome would suggest that the two phenomena are influenced by different mechanisms.

Why would KU be apparent in strategy effectiveness ratings (Figure 4), but not be reflected to the same extent in participants' postdictions and predictions? Note that the KU effects were barely registered in the JOLs that have been used in other research as the primary indicator of KU (Bieman-Copeland & Charness, 1994). This finding has now been widely replicated (Dunlosky & Hertzog, 2000; Hertzog et al., in press-a; Matvey et al., 2002). The different picture of KU that emerges from examining strategy effectiveness ratings versus metamemory judgments indicates that metamemory judgments should not be viewed as pure measures of declarative knowledge. The questionnaire ratings indicated that both younger and older adults learned that imagery is superior to rote repetition, and that blocked testing facilitated this KU process for younger adults. Yet the absolute accuracy of global predictions provided only qualified evidence of KU, and JOLs were largely insensitive to it. JOLs were higher for items studied with the imagery strategy, but the difference was far smaller than the actual differences in likelihood of recall. It appears that for JOLs, in particular, the strategy used is accessed when making the JOL, but it is apparently discounted in favor of other sources of information.

Individuals in both age groups lowered their global predictions after task exposure, consistent with prior research on KU (Dunlosky & Hertzog, 2000; Matvey et al., 2002). As in earlier work (Matvey et al., 2002), older adults responded to discrepancies between their postdictions and original predictions by lowering predictions for both imagery and repetition items, rather than differentially lowering their repetition ratings while maintaining higher predictions for imagery strategy items. Because recall of items studied with imagery tends to be higher than recall of items studied with repetition, downgrading of predictions for both types of strategies resulted in better absolute accuracy for repetition, but produced worse absolute accuracy for imagery. By contrast, younger adults in the blocking condition showed this general effect, but also less tendency to downgrade global predictions for imagery for List 2 (following their more accurate List 1 postdictions). Thus, global predictions showed weak evidence of KU for younger adults, and neither age group showed the magnitude of improvement in absolute accuracy of global predictions that might be expected given the changes in their questionnaire effectiveness ratings.

Our findings therefore strongly recommend the use of questionnaire ratings to examine KU, both because they have been shown to be sensitive to a manipulation that should influence KU, and because they directly query declarative knowledge, which is only one of the sources of information that may be accessed when individuals make metacognitive judgments. Ironically, Bieman-Copeland and Charness (1994) reported, but did not emphasize, that older adults' questionnaire responses seemed to indicate knowledge of differential cue effectiveness even when their JOLs showed no such effect. The present study brings us back full circle to this evidence, and

argues that the metacognitive judgments may in fact be less diagnostic of strategy knowledge than has been assumed.

The present study informs us about the potential ways in which individuals gain knowledge about strategy effectiveness during supervised learning, in which their strategic behavior is structured by the experimenter. One can wonder about the discovery process when strategic behavior is spontaneous and under participant control. Older adults are about as likely to use effective mediational strategies without instructions to do so (e.g., Dunlosky & Hertzog, 1998; Hertzog, Dunlosky, & Robinson, 2007). Do they also acquire strategy knowledge as part of this experience?

Another important question for future research is whether the improved knowledge gained from structured task experience affects strategy choices when individuals are free to choose any strategy they wish. It is possible that older adults would not use their new knowledge to select superior strategies to the same degree as younger adults. From a metacognitive perspective, the critical question, in general, is whether older adults use knowledge and on-line monitoring to adapt behaviors at study and test that optimize learning. Assessing older adults' flexibility in self-regulating learning is an important direction (e.g., Dunlosky & Connor, 1997; Hertzog & Dunlosky, 2004; Murphy, Schmitt, Caruso, & Sanders, 1987), one which requires considerably more evidence than is now available. Assessing knowledge and beliefs about strategies with task-specific questionnaires may be an important part of addressing that issue (Hertzog & Dunlosky, 2006), as it has been in studies of age differences in strategies for source monitoring (e.g., Henkel, Johnson, & De Leonardis, 1998; Johnson, 2006).

Original manuscript received 25 June 2007

Revised manuscript accepted 29 January 2008

First published online April 30, 2008

REFERENCES

- Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9*, 287–302.
- Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*, 249–257.
- Craik, F. I. M., Naveh-Benjamin, M., Ishaik, G., & Anderson, N. D. (2000). Divided attention during encoding and retrieval: Differential control effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1744–1749.
- Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging, 5*, 291–303.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory and Cognition, 25*, 691–700.
- Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597–607.

- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*, 462–474.
- 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.
- Finn, B., and Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 238–244.
- Henkel, L. A., Johnson, M. K. & De Leonardis, D. M., (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*, 251–268.
- 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. (2006). Using visual imagery as a mnemonic for verbal associative learning: Developmental and individual differences. In T. Vecchi & G. Bottini (Eds.), *Imagery and spatial cognition: Methods, models and cognitive assessment* (pp. 268–284). Amsterdam and Philadelphia, The Netherlands/USA: John Benjamins Publishers.
- Hertzog, C., Dunlosky, J., & Robinson, A. E. (2007). *Intellectual abilities and metacognitive beliefs influence spontaneous use of effective encoding strategies*. Unpublished manuscript.
- Hertzog, C., Saylor, L. L., Fleece, A. M., & Dixon, R. A. (1994). Metamemory and aging: Relations between predicted, actual and perceived memory task performance. *Aging and Cognition, 1*, 203–237.
- Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (in press-a). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology*.
- Hertzog, C., Price, J., & Dunlosky, J. (in press-b). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences*.
- Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*, 67–80.
- Johnson, M. K. (2006). Memory and reality. *American Psychologist, 61*, 760–771.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.
- 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.
- Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*, 83–97.
- Lovelace, E. A. (1990). Aging and metacognitions concerning memory function. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness, and interventions* (pp. 157–188). Amsterdam: North-Holland.
- Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge updating of cue effectiveness. *Psychology and Aging, 17*, 589–597.
- Murphy, M. D., Schmitt, F. A., Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging, 2*, 331–339.
- 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. <http://www.usf.edu/FreeAssociation/>.
- Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, *2*, 179–183.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.
- Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, *130*, 59–76.
- Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A micro-genetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.
- Visual Basic, Version 6.0 (1998). Microsoft Corporation.