Psychology and Aging
2000, Vol. 15, No. 3, 462–474

0882-7974/00/$5.00 DOI: 10.1037//0882-7974.15.3.462

Updating Knowledge About Encoding Strategies: A Componential Analysis of Learning About Strategy Effectiveness From Task Experience

John Dunlosky
University of North Carolina at Greensboro

Christopher Hertzog
Georgia Institute of Technology

Researchers have argued for age deficits in learning about the effects of encoding strategies from task experience, partly on the basis of absolute accuracy of metacognitive judgments. However, these findings could be attributed to factors other than age differences in learning. Forty older and 40 younger adults participated in 2 study–test trials in which they studied paired associates with imagery or repetition, predicted recall for the items, attempted recall, and postdicted recall. Recall was greater after imagery than repetition, yet this effect was not fully reflected by predictions made on Trial 1. Although both older and younger adults accurately postdicted recall from Trial 1, absolute accuracy of the predictions made on Trial 2 showed little improvement. By contrast, both age groups demonstrated increases in between-person correlations of predictions with recall, which is inconsistent with age deficits in knowledge updating. Thus, both older and younger adults had updated knowledge about the relative effects of the strategies, but such updating was not evident in the absolute accuracy of the predictions.

Metamemory pertains to one's cognition about his or her own memory and includes multiple components, such as knowledge about encoding strategies, on-line monitoring of study, and monitoring of test performance. Just as valid knowledge about how memory operates may support effective learning, a lack of such knowledge may constrain learning. Although research has shown that knowledge about memory increases during childhood (Schneider & Pressley, 1997), even adults do not necessarily know the relative benefits of many study strategies (e.g., Brigham & Pressley, 1988; Shaughnessy, 1981). Thus, an important goal for future research is to discover how individuals learn from experience about the effects of strategies, which we refer to as *knowledge updating*. In the present research, we investigated the degree to which adults learn about the relative effectiveness of two encoding strategies for associative learning. During two study–test trials, older and younger adults used both interactive imagery and rote repetition so they would have an opportunity to learn about their differential effects. We also collected indexes of multiple aspects of metacognition. Our main topics of interest included whether age-related deficits occurred in knowledge updating across trials and whether any differences there could be attributed to age-related deficits in metacognitive monitoring.

John Dunlosky, Department of Psychology, University of North Carolina at Greensboro; Christopher Hertzog, School of Psychology, Georgia Institute of Technology.

We thank Greg Matvey for comments on an earlier version of this article. This research was supported by National Institute on Aging Grant R01-AG13148.

Correspondence concerning this article should be addressed to John Dunlosky, P.O. Box 26164, Department of Psychology, University of North Carolina at Greensboro, Greensboro, North Carolina 27402–6164. Electronic mail may be sent to dunlosky@uncg.edu.

Age-Related Differences in Knowledge Updating

Limited evidence is available concerning age-related differences in knowledge updating. Bieman-Copland and Charness (1994) had adults study items during two study–test trials. During each study trial, participants made a judgment of learning (JOL) for each item, which is a prediction about the likelihood of recalling a recently studied item on an upcoming test. For both age groups, JOLs made on Trial 1 were about 15% greater for items with semantic (e.g., slippery–ice) cues than for items with rhyme cues (e.g., slice–ice), whereas recall was 40% better for semantic cues. Thus, the effects of these cues on Trial 1 were substantially larger on recall than on JOLs. A critical outcome concerning knowledge updating would be whether JOLs made on Trial 2 better reflected the differential effects of these cues. That is, after having experience with the task on Trial 1, would adults adjust JOLs to reflect the differences in cued recall, indicating that they had updated their knowledge about cue effectiveness? On the basis of JOLs made on Trial 2, Bieman-Copland and Charness (1994) concluded that "young subjects make differentiated adjustments in their predictions following recall experience. In contrast, old subjects make only global adjustments in the expectations, lowering their expectation for future performance across all conditions. . ." (p. 295). That is, younger adults updated their knowledge about the differential effects of the cues, whereas older adults did not.

Consider more closely the evidence that led to this conclusion. Younger adults' mean JOLs on Trial 2 increased for items with semantic cues and decreased for items with rhyme cues—that is, differentiated adjustments. However, for younger adults on Trial 1, mean recall for semantic cues was greater than mean JOLs for semantic cues, and mean recall for rhyme cues was less than mean JOLs for rhyme cues. Knowledge updating on Trial 2 could plausibly be manifested in a differentiated adjustment in JOLs. In contrast, for older adults on Trial 1, mean recall for both kinds of cue was lower than mean JOLs for those cues. Knowledge updating on Trial 2 would be more likely to be associated with a global shift in JOLs. The differences in patterns of change in JOL accuracy could be attributed to inconsistency between the age groups in the initial differences between cued recall and JOLs. Thus, perhaps both groups changed their JOLs in a way that reflected knowledge updating between trials.

Brigham and Pressley (1988) also reported evidence relevant to age differences in knowledge updating. They had participants study word–definition pairs using two strategies. Before practice with the strategies, older and younger adults on average had no strategy preference. After practice, younger adults more often chose the better strategy, whereas older adults did not. This outcome led Brigham and Pressley (1988) to conclude that older adults "appeared not to understand the relative utility of the strategies" (p. 255). However, the differential effect of the two strategies on recall was relatively weak for older adults, and hence they may have been less likely to discriminate between the effects of the strategies. If so, these age-related effects may have been determined more by age-related differences in the relative effectiveness of the strategies than by any deficits in knowledge updating.

In sum, although evidence reported in previous research suggests age deficits occur in knowledge updating, alternative interpretations are available. Thus, we explored age differences in knowledge updating with a design that addressed some of these concerns.

Framework of Knowledge Updating: Core Assumptions and Processes

To assist in explanation of our approach, we present a theoretical framework of knowledge updating (Figure 1). This framework represents how an individual learns about strategies by using them during multiple study–test trials.[1] Even though this multiple-trial procedure involves the kinds of experience individuals will often have when using a strategy, some cognitive and metacognitive components that comprise this framework are involved in knowledge updating that occurs in other tasks (Winne & Hadwin, 1998).

Any investigation of knowledge updating involves, at least implicitly, assumptions regarding how various cognitive and metacognitive components contribute to updating strategy knowledge and subsequently using this knowledge. Such investigations also must operationalize how knowledge about strategies can be measured, such as by metacognitive judgments. Our framework identifies four core assumptions that must hold if task experience is to result in knowledge updating, which is then manifested in people's subsequent predictions about the differential effectiveness of strategies. By making these assumptions explicit, one can identify conditions in which measures from an investigation are expected to demonstrate knowledge updating.

Effectiveness Assumption

The strategies an individual uses must be differentially effective. If not, one would not expect knowledge updating and hence would not expect experience with the strategies to influence people's subsequent predictions about them. Moreover, to evaluate whether age differences occur in knowledge updating, the relative effectiveness of the strategies should be similar for younger and older adults. As mentioned above, this assumption was not necessarily met in previous research.

Monitoring Assumption

An individual must accurately monitor the differential effectiveness of the strategies during the task, either while studying or while being tested. According to this assumption, knowledge updating cannot occur if an individual is unaware that learning or performance differs for items studied with different strategies. Age differences in monitoring could lead to age differences in knowledge updating.

Updating Assumption

Whereas the monitoring assumption refers to perceiving a relationship between strategies and their potential effects on learning, the updating assumption involves modifying knowledge about those strategies. Monitoring itself is not sufficient for knowledge updating. That is, for knowledge updating to occur, an individual must not only monitor differences among items but also attribute the differences among items to the strategies used to learn them and subsequently modify knowledge about those strategies. In Figure 1, knowledge updating is represented by links between monitoring and knowledge about strategies. The process of updating presumably involves a conscious inference about the efficacy of different strategies (but see Siegler & Shipley, 1995). Of course, knowledge updating itself refers to changes in people's memory about strategies, but such changes presumably rely on inferences about strategy effects. These inferences are based on experience with specific study or retrieval episodes and the accessibility of the products of monitoring from those episodes at the time the inference is made. Thus, age differences in knowledge updating could be caused by a lower probability for older adults to draw inferences about strategy effectiveness.

Utilization Assumption

Criterion performance following experience with the encoding strategies is determined (at least in part) by utilization of newly acquired knowledge about strategy effectiveness. As in Bieman-Copland and Charness (1994), the criterion tasks in the present research are predictions on Trial 2 made after individuals have had experience with the task on Trial 1 (see global-differentiated predictions and JOLs on Trial 2 in Figure 1). In principle, knowledge updating could occur yet not be utilized in making these predictions. Accordingly, age differences in the predictions (or any criterion task) reflect age differences in knowledge updating to the extent that one can assume age equivalence occurs in the utilization of strategy knowledge while making the predictions.

These core assumptions generate testable hypotheses concerning cognitive and metacognitive processes that may occur during

[1] Because this framework was developed to describe knowledge updating, we omitted less relevant details. For instance, updated knowledge may influence subsequent retrieval processes, and performance monitoring on one trial may directly influence strategy selection. Discovering whether these (and other) processes occur during learning is empirically tractable but not the focus of this research.

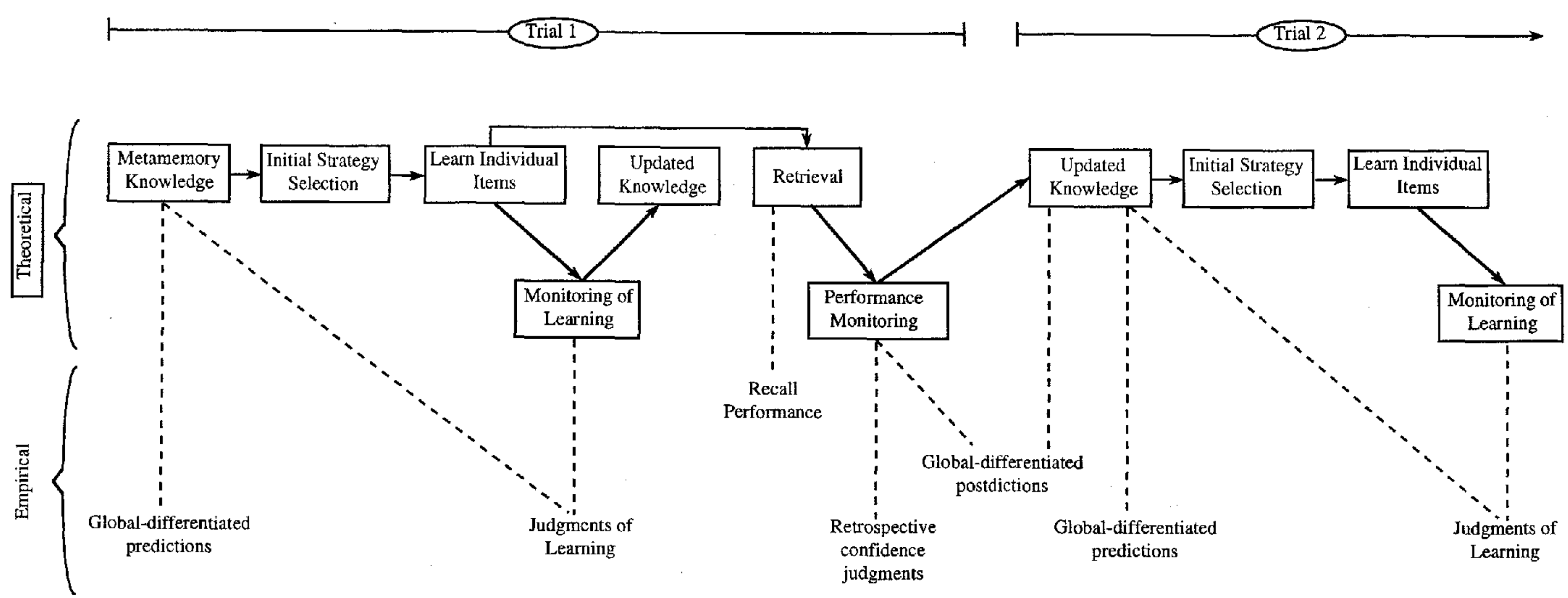

Figure 1. Theoretical framework of knowledge updating, illustrating the relations among metacognitive and cognitive components across two study–test trials. Theoretical constructs are presented in boxes in the top half of the figure, and empirical measures that tap those constructs are presented in the lower half. Dashed lines indicate the constructs presumably tapped by each measure. The entire set of processes that may occur during Trial 2 are not shown because they are not relevant to predictions evaluated in the present research.

multiple study–test trials. The present research was designed to satisfy the effectiveness assumption and also to provide data to test the assumptions in a comprehensive manner not attempted in previous research.

Does Predictive Accuracy Improve With Task Experience?

How does one measure whether task experience with strategies on one trial influences judgments on subsequent trials? As in Bieman-Copland and Charness (1994), our measures were based on how accurately people's metacognitive predictions indicate the effects of strategies on recall. However, we also extended their research by incorporating multiple metacognitive judgments that could provide more definitive evidence as to the source of any age differences in knowledge updating. As illustrated in Figure 1, each judgment taps different constructs related to the theoretical components of our framework.

To obtain these measures, we had adults study paired associates during two study–test trials. For half the items on a trial, participants were instructed to use imagery; for the other half, they were instructed to use repetition. During each trial, participants made two predictions: global-differentiated predictions and JOLs. *Global predictions*, made before study, involve predicting the number of items that will be recalled after study. Global predictions, as typically measured in the literature, query total recall for an entire list of items (e.g., Hertzog, Dixon, & Hultsch, 1990) and hence cannot differentiate between strategies used on different items in a within-subject design. Thus, we had participants make *global-differentiated* predictions, which involve predicting the percentage of items that will be recalled for each strategy (cf. Bruce, Coyne, & Botwinick, 1982). As illustrated in Figure 1, these predictions reflect in part knowledge about the effectiveness of each strategy. JOLs are a person's predictions of the likelihood of retrieving each item on an upcoming test. In contrast to global predictions, JOLs are made for each item and hence reflect monitoring during study of how a strategy influenced learning of each item.

Central outcomes of the present research concerned whether the accuracy of these two kinds of predictions increased across trials. On the first trial, these predictions were not expected to reflect the differential effects of imagery versus repetition (Dunlosky & Nelson, 1994). Provided that the core assumptions hold, an increase in predictive accuracy across trials provides evidence that participants updated knowledge about the effects of the strategies. Conversely, a failure by older adults to increase predictive accuracy might indicate an age-related deficit in knowledge updating.

Empirically Evaluating the Core Assumptions

If people failed to accurately change predictions after task experience, the design permitted an identification of the cognitive and metacognitive components responsible for this failure, including (but not limited to) knowledge updating. For instance, assume the JOLs did not reflect the differential effects of the strategies at either Trial 1 or Trial 2. In this case, task experience on Trial 1 presumably failed to improve the accuracy of JOLs. Why might such a failure occur? If the effectiveness assumption is valid, perhaps participants did not accurately monitor the differential effectiveness of the strategies during the test on Trial 1. This possibility can be empirically tested with a measure of performance monitoring. *Retrospective confidence (RC) judgments* are made immediately after testing recall of each item, asking a person to rate the likelihood that the response he or she provided to each item is correct. Poor accuracy of these judgments would disconfirm the monitoring assumption. We could conclude that people in general, or perhaps just older adults, are poor at monitoring performance (Bieman-Copland & Charness, 1994).

If the monitoring assumption holds, however, perhaps the participants did not infer that the strategies mediated the differences between the memory for individual items, violating the updating assumption. This assumption can be tested by examining global-differentiated *postdictions*, which are identical to the global-differentiated predictions, except that an individual estimates—immediately after completion of the entire Trial 1 test—how many items studied with each strategy had been recalled (Brigham & Pressley, 1988). Age differences in the accuracy of global-differentiated postdictions would implicate an age difference in inferences about strategy effectiveness from performance monitoring. Alternatively, if global-differentiated postdictions do reflect the differential effects of the strategies, then participants presumably did update knowledge about the strategies after test. Insufficient use of this updated knowledge when making JOLs would then be a plausible explanation for why JOL accuracy did not increase at Trial 2.

Method

Design, Materials, and Participants

The design was a 2 × 2 × 2 × 2 mixed factorial, with age (older vs. younger adults), trial (first or second study–test trial), strategy instructions (imagery vs. repetition), and order of global-differentiated judgments (imagery first or repetition first) comprising the four factors. Because the order of judgments was counterbalanced across participants and was a nuisance variable, it will not be discussed further.

Forty younger and 40 older adults participated. Younger adults (M age = 19.2 years, SD = 3.0) were students at the Georgia Institute of Technology and received course credit for participating. Older adults (M age = 69.7 years, SD = 5.7) were normal, community-dwelling adults recruited from both Atlanta, Georgia, and Greensboro, North Carolina areas. They received a nominal fee, and testing was conducted in campus laboratories.

Before the experiment began, a demographic questionnaire was administered. On a standard vocabulary test (maximum score of 36 points; Ekstrom, French, Harman, & Dermen, 1976), recognition performance was greater for older (M = 23.3, SD = 6.2) than for younger (M = 13.2, SD = 3.7) adults, $t(77)$ = 8.90. Older adults were well educated, as indexed by years of education (M = 15.3 years, SD = 2.4). Participants also reported using few medications and being in relatively good health, with mean ratings on a 4-point scale, ranging from 1 (excellent health) to 4 (poor health), being 1.6 (SD = 0.59) for younger and 1.4 (SD = 0.55) for older adults. This pattern is consistent with reports from previous research and indicates that older participants were relatively select, being well educated and healthy.

Critical items were 60 paired associates, with each item consisting of two concrete and relatively high frequency (concreteness [C] = 6.8; frequency [F] = 72.3; Paivio, Yuille, & Madigan, 1968) unrelated nouns (e.g., "dog–spoon"); the majority of words were one or two syllables (eight words had three syllables, and one had four syllables), with the mean number of letters per word equal to 5.4. Six extra items were used exclusively in a practice trial. During the experimental task, items were presented and responses collected on Macintosh computers.

Procedure

The task consisted of two study–test trials, with 30 different items being presented for study during each trial. Participants were instructed to study some items with interactive imagery and other items with rote repetition. Participants also made several metamemory judgments, each of which are described below. Prior to beginning the critical trials, participants (a) were given thorough instructions, which included descriptions and examples of both strategies, and explanations of the judgments, and (b) practiced studying and judging three items presented under each of the two strategy instructions.

Study. For each participant, half the critical items were chosen randomly and assigned to be presented during the first study–test trial, with the remaining items assigned during the second trial. The order of presentation of items during study was randomized. The presentation rate of items for study was 5 s for younger adults and 10 s for older adults. These rates of presentation were chosen because pilot data indicated that they would result in approximately equal recall for the two age groups, which was ideal for attaining comparable estimates of prediction accuracy for older and younger adults (Connor, Dunlosky, & Hertzog, 1997).

For each trial, half the items were randomly assigned to be studied by interactive imagery, and the other half assigned to study with rote repetition. Participants were explicitly instructed to use one of these two strategies for each item. One second prior to item presentation, a prompt appeared that instructed the participant to study the next item with either "Imagery" or "Repetition," reflecting the strategy assigned to that item. This prompt remained on the screen during item study. Previous research has shown that compliance to item-level strategy instructions is not perfect but that age differences in compliance are negligible (Dunlosky & Hertzog, 1998). Thus, measures of compliance were not collected.

Test. After items had been studied for a given trial, instructions for paired-associate recall were presented, and then participants attempted to recall the previously studied target responses. The order of items was randomized anew during the test trial. For each item, the stimulus was presented (e.g., "dog–?"), and the participant was instructed to type the word that was originally paired with the stimulus (e.g., "spoon"). Participants were given unlimited time to respond; omissions were not permitted. To minimize the role of spelling, we scored a response as correct if the first three letters matched the target response.

Metamemory judgments. Metamemory judgments occurred prior to study (global predictions), during study (judgments of learning), immediately after study but before the test (global predictions), during the test (RC judgments), and after the test (global postdictions). Each participant made the judgments at his or her own pace.

Immediately prior to beginning a study–test trial, participants made global-differentiated predictions. Participants were told that they would be instructed to study 15 items by means of imagery and 15 items by means of repetition. Each participant made two judgments, which indicated the percentage of items that he or she would recall after using imagery and after using repetition. In particular, they were asked to "type any number between 0 and 100 (inclusive) that corresponds to the percentage of pairs that you will study using Interactive Imagery [or Rote Repetition] that you think you will correctly recall." Global-differentiated predictions were also made immediately after study (see Connor et al., 1997). However, these predictions were not central to our present aims, and analyses indicated that they were essentially redundant with prestudy predictions. Thus, we did not consider these poststudy predictions further here. Instead, our analyses and discussion below focus on global-differentiated predictions made immediately prior to study.

Immediately after the offset of the presentation of each item for study, participants made a JOL. JOLs were prompted with only the stimulus of an item (e.g., if "dog–spoon" had been presented for study, the prompt would include "dog–?") 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?" (with answers ranging from 0 = *definitely won't recall*, 20 = *20% sure*, 40 = *40% sure*, 60 = *60% sure*, 80 = *80% sure*, and 100 = *definitely will recall*).

During the test, participants made a RC judgment immediately after providing a response to an item. The confidence judgments were prompted with the query, "How confident are you that the answer you just gave is correct?" (with answers ranging from 0 = *definitely was not correct*, 20 = *20% sure it was correct*, 40 = *40% sure it was correct*, 60 = *60% sure it was correct*, 80 = *80% sure it was correct*, and 100 = *definitely was correct*).

Finally, immediately after all items had been presented for paired-associate recall for a given trial, participants made global-differentiated postdictions. Namely, participants were asked to postdict the percentage of items that were correctly recalled. Postdictions were made separately for items studied under imagery instructions and for those studied under repetition instructions, using the same prompt for global-differentiated predictions (see above) but referring to the percentage of words correctly recalled on the test.

Results

Recall Performance

Although predictive accuracy is most relevant to evaluating knowledge updating, we present recall performance first. If the strategies did not differ in level of recall for both age groups, which would violate the effectiveness assumption, then further analyses of knowledge updating would not be meaningful. Moreover, an ideal test of age differences in knowledge updating involves age equivalence in the magnitude of strategy effects on recall, so that both age groups have similar opportunity to observe performance differences caused by the two strategies. Thus, we gave older adults more study time in an attempt to make recall performance similar for the two age groups (but see Kausler, 1994). Means of individuals' proportions of correct recall are reported in Table 1.

A 2 (age: younger vs. older adults) × 2 (trial: Trial 1 vs. Trial 2) × 2 (instructions: imagery vs. repetition) analysis of variance (ANOVA) revealed a main effect for trial, $F(1, 78) = 11.3$, $MSE = 0.01$, and for instructions, $F(1, 78) = 56.9$, $MSE = 0.05$. The main effect of age and all interactions involving age were not reliable ($Fs < 1.40$). As evident from inspection of Table 1, the benefits of imagery (over repetition) instructions were nearly identical for older and younger adults. Thus, the different presentation rates succeeded in making older and younger adults comparable in strategy effectiveness. Hence, any age-related differences in the accuracy of metamemory judgments are unlikely to be due to age deficits in recall performance.

Measures of Predictive Accuracy

Predictive accuracy is not a unitary construct, but rather consists of relative accuracy and absolute accuracy. Accuracy may also be calculated within an individual or between different individuals, which results in multiple measures of predictive accuracy. In total, we report five measures to evaluate whether the accuracy of predictions (JOLs and global-differentiated predictions) increased across trials. Three measures were calculated for each individual: relative accuracy of JOLs, absolute accuracy of JOLs, and absolute

Table 1
Magnitude of Recall, Predictions, and Postdictions

	Trial 1				Trial 2			
	Imagery		Repetition		Imagery		Repetition	
Measure	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)
			Younger adults					
Recall	56	(4.0)	35	(3.5)	58	(4.3)	43	(3.7)
Global predictions	43	(4.1)	42	(2.7)	42	(4.2)	30	(3.2)
JOLs	47	(3.4)	44	(3.2)	44	(4.1)	36	(3.2)
Confidence	55	(3.8)	34	(3.0)	59	(4.2)	43	(3.4)
Global postdictions	36	(4.5)	22	(3.0)	41	(4.6)	28	(3.6)
			Older adults					
Recall	50	(4.1)	31	(3.4)	53	(4.4)	35	(4.0)
Global predictions	43	(3.3)	36	(3.1)	30	(2.8)	21	(3.0)
JOLs	38	(2.7)	31	(2.7)	33	(3.0)	25	(2.7)
Confidence	48	(3.6)	29	(3.1)	51	(4.2)	34	(3.8)
Global postdictions	31	(3.6)	19	(2.5)	35	(3.8)	20	(2.8)

Note. All global judgments refer to global-differentiated judgments. JOLs = judgments of learning; Confidence = retrospective confidence judgments made for responses during paired-associate recall; Imagery = items studied under imagery instructions; Repetition = items studied under repetition instructions. Entries are mean percentages (and standard errors) for the measures.

accuracy of global-differentiated predictions. Two measures involved between-person correlations of judgments with recall. Because these measures are not merely different indicators of a single hypothetical construct, one measure may show an increase across trials when others do not (for further discussion and evidence of empirical dissociations among them, see Connor et al., 1997; Hertzog & Hultsch, 2000; Koriat, 1997). For example, as compared with predictions on Trial 1, individuals' predictions on Trial 2 may align better with individual differences in recall, whereas the discrepancy between mean predictions and mean recall may increase. In the former case, the increase in the accuracy of individuals' predictions across trials would be revealed by increases in between-person correlations of predictions and recall. In the latter case, the decrease in accuracy would be revealed by measures of absolute accuracy. In terms of psychological processes, these outcomes would indicate that some individuals adjusted their predictions (relative to others' predictions) to compensate for initially underestimating (or overestimating) recall on Trial 1 but also that (on average) the magnitude of the predictions diverged from the absolute effects of study on recall. Given the possibility of such dissociations, we report all five measures of accuracy, which are described in detail below.

Accuracy of Judgments of Learning and Global-Differentiated Predictions

Within-individual analyses. Two kinds of within-individual measures, relative accuracy and absolute accuracy, were calculated to assess whether the accuracy of people's JOLs increased across trials. Relative accuracy was measured by a Goodman-Kruskal gamma correlation between an individual's JOLs and recall performance. Gamma correlations are not influenced by the absolute magnitude of JOLs or recall but instead assess the accuracy of JOLs at predicting the recall of one item relative to another. The mean gamma correlations between each individual's JOLs and recall are reported in Table 2. Separate analyses were conducted

Table 2
Relative Accuracy of Judgments of Learning

	Trial 1						Trial 2					
	Imagery		Repetition		Aggregate		Imagery		Repetition		Aggregate	
Age group	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)
Younger adults	.41	(.07)	.38	(.09)	.45	(.04)	.36	(.09)	.44	(.08)	.43	(.07)
Older adults	.28	(.10)	.37	(.10)	.48	(.06)	.31	(.10)	.39	(.12)	.48	(.12)

Note. Entries are means (and standard errors) of individuals' Goodman-Kruskal gamma correlations between judgments of learning and recall performance. Correlations were calculated separately for items studied under imagery instructions (Imagery) and under repetition instructions (Repetition) as well as aggregated across all items (Aggregate).

Table 3
Difference Scores Between Predictions and Recall

Age group	Trial 1				Trial 2			
	Imagery		Repetition		Imagery		Repetition	
	M	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)
Judgments of learning								
Younger adults	−8.4	(4.6)	8.9	(4.6)	−13.7	(4.2)	−7.4	(4.4)
Older adults	−11.9	(4.4)	0.7	(3.9)	−19.9	(4.0)	−10.4	(3.3)
Global-differentiated predictions								
Younger adults	−12.4	(4.7)	7.2	(4.8)	−15.4	(3.9)	−12.8	(4.0)
Older adults	−5.1	(4.7)	5.8	(4.2)	−21.5	(3.6)	−13.6	(3.5)

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 studied under imagery instructions; Repetition = items studied under repetition instructions.

for accuracy within strategy instructions (labeled *Imagery* and *Repetition* in Table 2) and for accuracy aggregated across strategy instructions (labeled *Aggregate* in Table 2).

Accuracy within strategy instructions was analyzed using a 2 (age: younger vs. older adults) × 2 (trial: Trial 1 vs. Trial 2) × 2 (instructions: imagery vs. repetition) ANOVA. The main effect of age approached significance, $F(1, 50) = 3.93$, $MSE = .34$, $p = .053$, with accuracy being somewhat lower for older than younger adults. The main effect of instruction and interactions involving instructions were not reliable ($Fs < 2.1$, $MSEs < .34$) and the main effect of trial and all interactions involving trial were not reliable ($Fs < 1.2$, $MSEs < .32$). Thus, relative JOL accuracy for items studied under a particular strategy did not improve across trials, suggesting that individuals did not learn from experience about item characteristics that made items more or less likely to be recalled.

In contrast to accuracy calculated for items studied within a strategy instruction, gamma correlations for all items, aggregating across strategy instructions, can be influenced by the degree to which JOLs differ for items studied under imagery versus repetition instructions. Thus, if people's JOLs become sensitive to the effects of the two strategies after task experience, the aggregate gamma correlation should increase across trials. However, a 2 (age: older vs. younger adults) × 2 (trial: Trial 1 vs. Trial 2) ANOVA revealed neither reliable main effects nor an Age × Trial interaction ($Fs < 1.0$, $MSEs < .20$). Note, however, that the sample means suggest that older adults' aggregate accuracy was higher than the JOL accuracy for each of the two strategies at both trials, indicating some influence of strategy instructions on JOLs. This was not the case for younger adults.

Even though relative accuracy was not influenced by task experience, participants may have learned about the magnitude of differences between the two strategies, which would be reflected in absolute accuracy of the predictions. For absolute accuracy, the overall level of predicted recall for a group of items (e.g., items slated with instructions to use imagery) is compared with the actual level of recall for that group of items (for concerns regarding interpretations of measures of absolute accuracy, see Keren, 1991; Wallsten, 1996). In contrast to relative accuracy, improvements in absolute accuracy can be measured both for JOLs and for global-differentiated judgments. We computed the difference between each individual's mean JOLs (or global predictions) and percent recall. Increases in absolute accuracy across trials would suggest that people gained knowledge about the absolute magnitude of recall for the two strategies.

Absolute accuracy is reported in Table 3.[2] Positive values indicate overestimations of recall, and negative values indicate underestimations. Because the magnitude of predictions themselves are a key component of absolute accuracy, we also report the sample means of these predictions in Table 1.

As is evident in Table 3, absolute accuracy of JOLs decreased across trials, $F(1, 78) = 23.0$, $MSE = 358.2$, with participants having a greater tendency to underestimate performance on Trial 2. Absolute accuracy was worse for items studied under imagery than for those studied under repetition instructions, $F(1, 78) = 40.2$, $MSE = 260.4$. Most important, the main effect of age and all interactions involving age were not reliable ($Fs < 2.70$).

The absolute accuracy of global predictions for each strategy also decreased across trials, $F(1, 76) = 39.9$, $MSE = 463.3$, and again was reliably worse for items studied under imagery than for those studied under repetition instructions, $F(1, 76) = 23.1$, $MSE = 381.3$. Although the sample means suggested that older adults had lower absolute accuracy than younger adults, the main effect of age and the interactions involving age were not reliable ($Fs < 3.2$). The most critical outcome, however, was that task experience reliably reduced absolute accuracy of global predictions for both age groups.

Between-person analyses. Another measure of accuracy is the degree to which individual differences in predictions and recall are correlated, using standard indexes of correlation such as Pearson's *r*. Between-person correlations of predictions and performance have been found to increase over trials, even when absolute accuracy did not improve (Hertzog, Saylor, Fleece, & Dixon, 1994). Correlations were computed for individuals' mean JOLs or global predictions with their recall, separately for each strategy. Increases in between-person accuracy across trials would indicate that indi-

[2] Conclusions involving age-related differences in these and other metacognitive variables reported later were identical for signed values and for unsigned values, so we present only signed values.

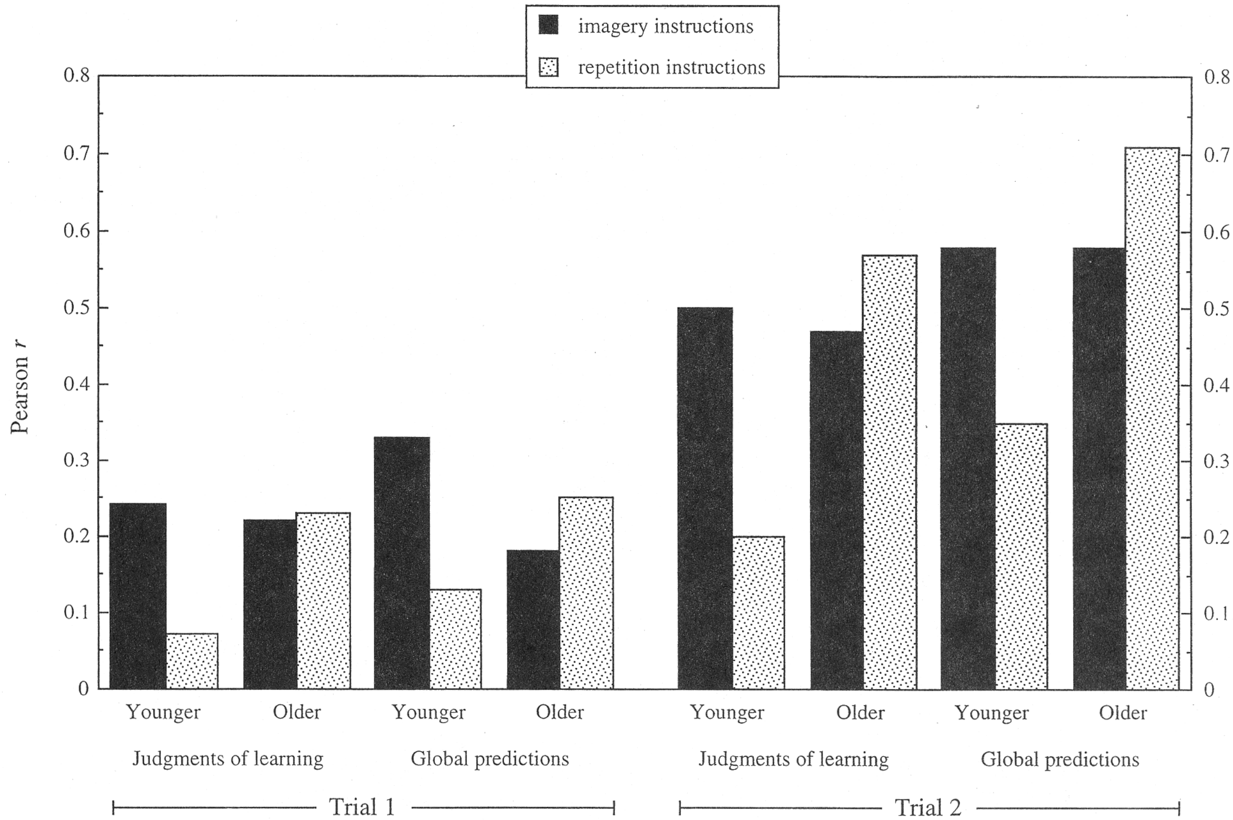

Figure 2. Pearson *r* correlations calculated between (a) individuals' mean judgments of learning and mean recall performance and (b) individuals' global-differentiated predictions and mean recall performance. For each trial, correlations were calculated separately for items studied under imagery instructions and under repetition instructions.

viduals' adjustments of either prediction were consistent with individual differences in recall.

Figure 2 reports the Pearson *r* correlations of the magnitude of individuals' mean JOLs (and global predictions) with recall performance. Two general outcomes are evident in Figure 2. The accuracy of the predictions improved with task experience, with each of eight correlations being greater on Trial 2 than on Trial 1, $p = .008$, by means of a sign test. As important, age-related differences were minimal (or favored older adults) in the increase in correlations across trials. These observations were generally supported by a series of dependent samples tests of increases in correlations over trials, which were conducted separately for both age groups. These tests were conducted using LISREL 8 to compute maximum likelihood tests of the equality of related correlations (Hertzog et al., 1994). When the correlations between JOLs and recall at Trial 1 were constrained to be equal to the same correlations at Trial 2, the resulting model provided a poor fit to the correlations of older adults, $\chi^2(2) = 9.0, p < .05$, but did not produce a loss of fit for younger adults, $\chi^2(2) = 3.88, p < .14$. For older adults, simple effects tests indicated reliable increases in correlations across trials for JOLs made for imagery items ($z = 2.09$) and for repetition items ($z = 2.60$). For younger adults, simple effects tests indicated an increase for imagery items ($z = 1.97$), whereas the increase for repetition items was not reliable ($z = 0.76$). When correlations between global-differentiated predictions and recall were constrained to be equivalent across Trial 1 and Trial 2, the resulting loss of fit was significant for older adults, $\chi^2(2) = 13.2, p < .05$, but not for younger adults, $\chi^2(2) = 4.62, p < .10$. For older adults, simple effects tests indicated reliable increases in correlations across trials for global predictions made for imagery items ($z = 2.36$) and for repetition items ($z = 3.05$). For younger adults, the simple-effects test approached reliability for imagery items ($z = 1.90$) but not for repetition items ($z = 1.42$).

In sum, between-person correlations increased from Trial 1 to Trial 2, indicating that individuals changed their predictions across trials in a way that more accurately reflected individual differences in recall. These results are inconsistent with the hypothesis of age deficits in knowledge updating.

Accuracy of Retrospective Confidence Judgments and Global-Differentiated Postdictions

Younger and older adults did not demonstrate much learning about strategies in their Trial 2 predictions, with only between-person correlations increasing across trials. In this section, we sought to localize the source of the difficulties by examining the accuracy of RC judgments and global-differentiated postdictions.

Table 4
Difference Scores Between Postdictions and Recall

Age group	Trial 1				Trial 2			
	Imagery		Repetition		Imagery		Repetition	
	M	(*SE*)	*M*	(*SE*)	*M*	(*SE*)	*M*	(*SE*)
Retrospective confidence judgments								
Younger adults	−0.80	(2.0)	−0.83	(1.4)	0.90	(1.7)	−0.13	(1.5)
Older adults	−2.43	(2.3)	−1.40	(2.0)	−2.07	(2.3)	−1.13	(1.9)
Global-differentiated postdictions								
Younger adults	−20.0	(3.6)	−12.4	(2.6)	−16.7	(3.5)	−15.3	(3.3)
Older adults	−19.7	(3.5)	−12.0	(2.6)	−18.4	(3.1)	−15.4	(2.6)

Note. Entries are means (and standard errors) of individuals' difference scores between postdictions (either retrospective confidence judgments or global-differentiated postdictions) and recall. Imagery = items studied under imagery instructions; Repetition = items studied under repetition instructions.

Within-individual analyses. To measure relative accuracy of the RC judgments, we calculated a gamma correlation between each individual's RC judgments and recall. On all trials, relative accuracy was high. For Trial 1, mean accuracy was .98 for younger adults and .94 for older adults (*Mdn* = .99 for both groups). For Trial 2, mean accuracy was .99 for younger adults and .96 for older adults (*Mdn* = 1.00 and .98 for younger and older adults, respectively). The median accuracy for both groups was also 1.0 for items studied under either strategy instruction. Thus, accuracy here was essentially perfect, which obviated the need for inferential statistics.

In Table 4, we report the measures of absolute accuracy, the signed difference scores, separately for RC judgments and for global-differentiated postdictions. The sample means of these postdictions are included in Table 1.

Absolute accuracy of RC judgments was high—all main effects and interactions were not statistically reliable ($Fs < 1.0$). The lack of age effects is also evident in the mean values of recall and RC judgments reported in Table 1.

However, global-differentiated postdictions consistently underestimated performance, with all values being reliably less than zero ($ts > 4.5$). The Instructions × Trial interaction was reliable, $F(1, 78) = 4.22$, $MSE = 143.5$, indicating absolute accuracy was worse for imagery items than for repetition items on Trial 1, with accuracy on Trial 2 increasing somewhat for imagery items but decreasing for repetition items. The main effect of age and all interactions involving age were not reliable ($Fs < 1.0$). Such poor absolute accuracy contrasts the high level of absolute accuracy for RC judgments. Even so, as shown in Table 1, global postdictions discriminated between imagery and repetition items about equally for both age groups—the effects for instruction and trial were reliable ($Fs > 6.0$, $MSEs < 357.0$), whereas the effect of age and interactions involving age were not reliable ($Fs < 1.80$, $MSEs < 1460.1$).

Performance monitoring during recall was highly accurate, which empirically validates the monitoring assumption. Thus, the poor absolute accuracy of the Trial 2 predictions cannot be attributed to inaccurate performance monitoring. Although global postdictions substantially underestimated recall, they did discriminate between imagery items and repetition items. Both older and younger adults appeared to infer after Trial 1 recall that the strategies were differentially effective, consistent with the updating assumption.

Between-person analyses. Even though the global-differentiated postdictions substantially underestimated performance, they could accurately capture individual differences in recall. Accordingly, we computed Pearson *r* correlations between the magnitude of individuals' postdictions and recall, which are reported in Table 5. Between-person correlations were high for both kinds of postdiction, suggesting a source of the increased predictive accuracy on Trial 2 is the information tapped by Trial 1 postdictions. We explore this possibility further in the path analyses described below.

Path Analyses: What Accounts for Individual Differences in Predictions on Trial 2?

The correlations just reported suggest that increases in predictive accuracy on Trial 2 may have been produced by the high accuracy of postdictions made after Trial 1 recall. This pattern of results could represent updating of knowledge about the strategies gained from task experience. We conducted path analyses to test

Table 5
Pearson r Correlation Coefficients of Individuals' Postdictions and Recall Performance

Age group	Trial 1		Trial 2	
	Imagery	Repetition	Imagery	Repetition
Retrospective confidence judgments				
Younger adults	.86	.93	.92	.92
Older adults	.83	.82	.86	.88
Global-differentiated postdictions				
Younger adults	.64	.69	.68	.58
Older adults	.57	.62	.73	.77

Note. Entries are means of individuals' difference scores between postdictions (either retrospective confidence judgments or global-differentiated postdictions) and recall. Imagery = items studied under imagery instructions; Repetition = items studied under repetition instructions.

this hypothesis. One model tested updating in JOLs, using mean JOLs on Trial 1, mean recall on Trial 1, Trial 1 RC judgments after recall, global-differentiated postdictions on Trial 1, and mean JOLs on Trial 2 as the relevant variables. The replicated measures for both imagery and repetition items were included. A second model tested updating in global-differentiated predictions, substituting global-differentiated predictions for JOLs. We pooled the data for both age groups to increase statistical power.

The models provided critical tests of several aspects of our theoretical framework. In particular, the monitoring and knowledge updating assumptions stipulate that if accurate performance monitoring occurs, it can then be subsequently used to update strategy knowledge, which in turn influences predictions on Trial 2. Hence, a critical question was, do RC judgments and global postdictions on Trial 1 mediate relations of Trial 1 recall to subsequent predictions on Trial 2? For both dependent variables (JOLs and global predictions on Trial 2), the models we estimated supported this mediation hypothesis and also validated other aspects of the theoretical framework.

The basic specification for both models incorporated a flow of causation from early to late in the sequence of variables within an experimental condition (either imagery or repetition). Relations between the two conditions were modeled only as correlated residuals between the mirrored imagery and repetition variables, under the assumption that similar variables share specific sources of variance, which were not our primary focus of interest. For instance, residuals for Trial 2 global imagery and repetition predictions were allowed to correlate, assuming that there would be consistency in prediction behavior for the two kinds of items at Trial 2 that was not a function of earlier predictions or postdictions.

This model specification worked well for both global predictions and JOLs. The fits of the model were relatively good: for JOLs, $\chi^2(28, N = 78) = 56.01, p = .001$, Bentler's (1990) CFI = .961, standardized root mean square residual (SRMR) = .095; and for global-differentiated predictions, $\chi^2(28, N = 78) = 46.74, p = .015$, CFI = .973, SRMR = .108. However, LISREL Modification Indices suggested an additional relationship between repetition RC judgments and imagery postdictions at Trial 1. Although we had not expected this pattern, it is interpretable. This path suggests that participants do not have perfect recollection of the strategy that had been used to study items, so that poor performance for the repetition items compromised people's perceived effectiveness of imagery. Adding this path improved the fit for both models: for JOLs, $\chi^2(27, N = 78) = 44.14, p = .02$, CFI = .976, SRMR = .069; and for global predictions, $\chi^2(27, N = 78) = 34.72, p = .15$, CFI = .989, SRMR = .053. Hence, the extra path was retained in both cases.

The standardized parameter estimates are presented in Figure 3 (JOLs) and Figure 4 (global predictions). RC judgments for imagery items and repetition items were highly related to prior recall performance, independent of Trial 1 predictions. This outcome indicates accurate performance monitoring during test. Furthermore, RC judgments significantly influenced global-differentiated postdictions, suggesting knowledge updating about strategies after the completion of recall. Given that imagery postdictions were influenced by both imagery and repetition RC judgments, inferences about strategy effectiveness were not entirely accurate, suggesting some forgetting of the strategies used to study items.

Most critically, global postdictions had paths to JOLs on Trial 2 independent of JOLs on Trial 1, indicating an influence of knowledge updating on the accuracy of JOLs made on Trial 2 (Figure 3).

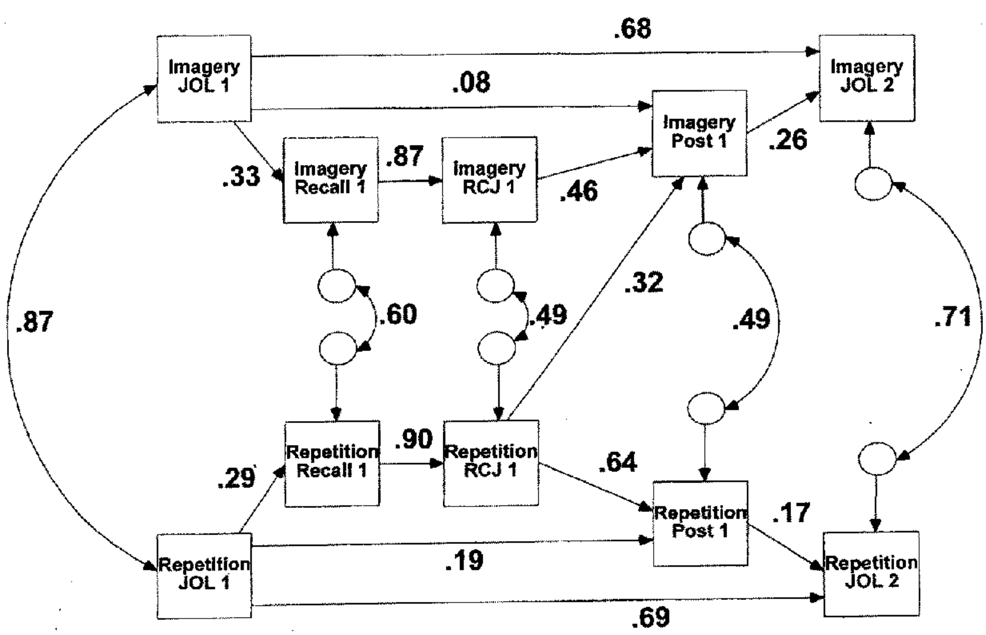

Figure 3. Path model accounting for individual differences in judgments of learning (JOL) on Trial 2 (JOL 2) with variables on Trial 1 (items with the suffix "1") relevant to on-line monitoring during study, performance monitoring, and knowledge updating. Imagery = items studied under imagery instructions; Repetition = items studied under repetition instructions; RCJ = retrospective confidence judgments; Post = global-differentiated postdictions.

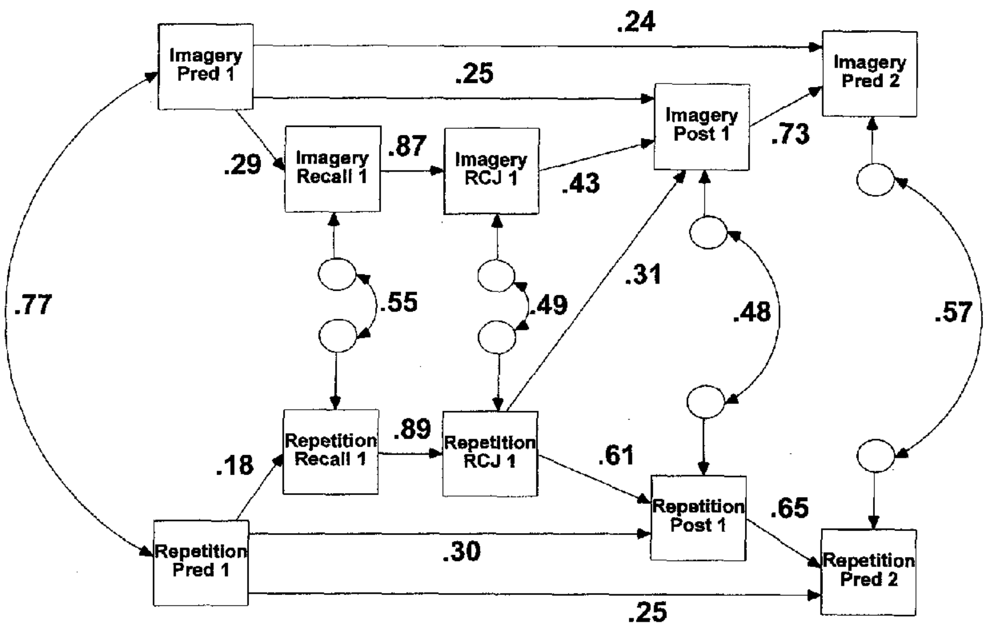

Figure 4. Path model accounting for individual differences in global-differentiated predictions (Pred) on Trial 2 (Pred 2) with variables on Trial 1 (items with the suffix "1") relevant to a priori knowledge, performance monitoring, and knowledge updating. Imagery = items studied under imagery instructions; Repetition = items studied under repetition instructions; RCJ = retrospective confidence judgments; Post = global-differentiated postdictions.

A parallel pattern emerged for global predictions (Figure 4), but the magnitude of these effects differed. In particular, JOLs showed relatively high stability from Trial 1 to Trial 2, along with weaker paths from prior global postdictions to mean JOLs on Trial 2. In contrast, global predictions on Trial 2 were strongly related to Trial 1 global postdictions and were modestly related to Trial 1 global predictions.

The mediation hypothesis was further evaluated by estimating alternative models. Adding paths from recall to predictions, independent of the postdictions, did not enhance model fit and still manifested large effects of postdictions on Trial 2 global predictions (or JOLs). We also disconfirmed the hypothesis that postdictions had no direct influence on Trial 2 predictions. In alternative models, recall was allowed to influence both RC judgments and Trial 2 global predictions (or JOLs), but the paths from postdictions to Trial 2 predictions (or JOLs) were eliminated. Thus, recall was modeled as influencing predictive behavior at Trial 2, without mediation through the intervening metacognitive judgments. Forcing the updating effect to be independent of the global postdictions degraded the fit for JOLs, $\chi^2(27, N = 78) = 50.20$, $p < .01$, CFI = .968, SRMR = .079, and for global predictions, $\chi^2(27, N = 78) = 116.44$, $p < .001$, CFI = .870, SRMR = .112. Given the weaker effects of postdictions on JOLs, the differences in loss of fit for the two models was not surprising. The important point is that the mediation hypothesis (i.e., performance monitoring drives updating, which in turn drives subsequent predictions) provided the best fit to the data in both cases.

Discussion

The present method supported three main conclusions: (a) Aging leaves the ability to monitor memory relatively intact; (b) predictive accuracy improved across trials in the between-person analyses despite little improvement in accuracy in several other measures; and (c) aging has little influence on updating knowledge about the differential effects of two encoding strategies. The first conclusion was apparent in every measure of accuracy: Even though the judgments were seldom perfectly accurate, accuracy was typically above chance (e.g., γs > 0 for relative accuracy) and not different for the two age groups. Because this phenomenon has been discussed in detail elsewhere (e.g., Connor et al., 1997; Hertzog & Hultsch, 2000), we do not discuss this conclusion further. Instead, we focus on the latter two conclusions, which directly pertain to learning about strategy effectiveness.

Explaining Success and Failure in Knowledge Updating

The framework for metacognitive judgments shown in Figure 1 allows one to evaluate alternative explanations of success and failure in knowledge updating. Concerning the former, Pearson *r* correlations increased across trials for both age groups (Figure 2). Outcomes from path analyses were consistent with the interpretation that metacognitive monitoring leads to knowledge updating, which in turn influences subsequent predictions. That is, associative recall for each strategy was not the best predictor of individuals' adjustments to Trial 2 predictions. Instead, the relationship of Trial 1 recall to Trial 2 predictions was mediated by performance monitoring and subsequent knowledge updating about the two strategies, as reflected in the global-differentiated postdictions.

In contrast to improvements in the between-person analyses of accuracy, other measures of accuracy either decreased or showed little improvement across trials. The magnitude of global-differentiated postdictions made on Trial 1 did discriminate between the two strategies (Table 1), suggesting that older and younger adults had gained some knowledge about the relative effects of the strategies. Thus, the lack of increase in accuracy across trials may have arisen because individuals did not appropriately utilize knowledge manifested after Trial 1 in making predictions on Trial 2. Of course, the present method did not provide a direct measure of utilization, which leaves open the possibility that other factors (e.g., forgetting) may also constrain accuracy of the Trial 2 predictions.

One particularly important outcome of the path analyses is the discrepancy in knowledge updating between global-differentiated predictions and JOLs. Global postdictions were strongly related to Trial 2 global predictions, and the consistency of global-differentiated predictions across trials was relatively low (cf. Connor et al., 1997; Lachman, Steinberg, & Trotter, 1987). In contrast, JOLs showed much higher consistency of individual differences across trials, even though global postdictions were related to JOLs and mediated the weaker updating effects from Trial 1 recall. One possible explanation is that updated knowledge about strategy effectiveness was less likely to be accessed when making JOLs during Trial 2 than when making global predictions immediately before study. Another explanation is that JOLs are affected by decision criteria for scaling confidence in recall of individual items that do not influence global predictions and that are not sensitive to knowledge about strategies. A critical assumption in the work by Bieman-Copland and Charness (1994) is that knowledge about cue effectiveness will be manifested in JOLs. In both that research and the present research, however, JOLs were obtained on scales with limited response options. These scales may have constrained the opportunity for knowledge to influence JOLs.

The results from the path analyses suggest an explanation for the underestimation of recall by Trial 1 postdictions. Recall for repetition items had an effect on both repetition and imagery postdictions. Individuals with low levels of recall after repetition apparently downgraded their imagery postdictions as well. This outcome suggests participants partly confused (either during recall or while making postdictions) the strategies used for studying individual items. Limited accessibility to strategies used for different items could also explain the poorer absolute accuracy for imagery items, relative to repetition items.

Finally, a comparison of the method used here with the method used by Bieman-Copland and Charness (1994) provides insight into why adults may have been limited in updating. Whereas recall was about 20% greater after imagery than repetition, Bieman-Copland and Charness (1994) found that recall for items with semantic cues (vs. items with rhyme cues) was about 40% higher for older adults and about 60% higher for younger adults. Perhaps larger experimental effects on recall increase the likelihood of updating (cf. Kruschke & Johansen, 1999). Inferences about the effects of the strategies may be more accurate when the strategies have a substantial influence on recall.

Aging and Knowledge Updating

Global-differentiated predictions and JOLs aligned better with individual differences in recall on Trial 2 than on Trial 1. These increases in accuracy were as large for older adults as they were for younger adults, demonstrating that aging does not necessarily reduce learning about strategy effects. Although these results are inconsistent with those from previous research, a number of differences between investigations may account for the apparently inconsistent results.

First, between-person analyses were not reported in previous research, so perhaps similar evidence of updating would have been obtained with this measure of accuracy. Instead, both Bieman-Copland and Charness (1994) and Brigham and Pressley (1988) focused on absolute accuracy, a measure of accuracy that may be biased by overall levels of memory performance. Given that age differences were evident in recall in these investigations, analyses of absolute accuracy reported there could be attributed to deficits in memory rather than to deficits in metamemory (Connor et al., 1997). We sidestepped this potential problem by giving older adults longer presentation rates to minimize age differences in recall. Nevertheless, absolute accuracy of JOLs and global predictions did not improve significantly across trials for either age group.[3]

A second difference pertains to the use of postdictions, which were not collected by Bieman-Copland and Charness (1994). Increases in predictive accuracy reported in the present research may be attributable to forcing participants to make postdictions, which may draw attention to the relative effectiveness of the strategies. That is, requiring postdictions after Trial 1 recall may have elicited updating that would not have been observed otherwise. This effect in turn could have increased correlations of Trial 2 predictions and recall. Thus, we had 26 younger adults and 30 older adults (from the same population) participate in the same experiment, except that JOLs were the only judgment made. Within-individual analyses of relative and absolute JOL accuracy were consistent with outcomes reported above, with negligible age effects and no systematic increases in accuracy across trials. Similar to outcomes presented in Figure 2, accuracy increased across trials. For younger adults, Pearson r correlations between JOLs and recall for imagery items were .07 for Trial 1 and .20 for Trial 2. For repetition items, the corresponding correlations were .23 for Trial 1 and .46 for Trial 2. For older adults, Pearson r correlations between JOLs and recall for imagery items were .42 for Trial 1 and .65 for Trial 2. For repetition items, the corresponding correlations were .27 for Trial 1 and .54 for Trial 2. Thus, increases in between-person accuracy of JOLs occurred even when participants did not make other metacognitive judgments.

Both Bieman-Copland and Charness (1994) and Brigham and Pressley (1988) concluded that age deficits reported in their research were partly attributable to age deficits in monitoring. The

[3] In contrast to the measures of absolute accuracy, the sample means (for global predictions and JOLs) suggested some age-related differences in updating. For instance, younger adults' mean predictions did not differ for imagery versus repetition items on Trial 1, yet their mean predictions were higher for imagery items than for repetition items on Trial 2. Although the same pattern is not apparent for older adults, this outcome should be interpreted cautiously for three reasons: (a) The critical three-way interaction involving age was not statistically reliable for global predictions or for JOLs ($Fs < 2.20$); (b) the interactions apparent in the mean values (Table 1) were much less pronounced in group medians for these variables, showing nearly identical (and quite minimal) updating for older and younger adults; and (c) older adults' sample means already differentiated somewhat between the strategies on Trial 1, rendering interpretation of the observed age differences in updating relatively ambiguous.

results presented here are inconsistent with this explanation, because they demonstrate negligible age-related differences in performance monitoring. Part of this discrepancy concerns how monitoring was operationalized by different investigators. We distinguished on-line monitoring of performance, as measured by RC judgments, from inferential processes about performance after recall, as partly measured by global postdictions. Outcomes from the present research support this distinction. For instance, RC judgments showed near-perfect accuracy, whereas global postdictions showed relatively poor absolute accuracy for both age groups. In the previous research, both kinds of processes were subsumed under the construct of monitoring.

Finally, Bieman-Copland and Charness (1994) concluded that younger adults differentially adjust predictions on Trial 2, "whereas older adults are only able to make much less differentiated, global adjustments in their [JOLs]" (p. 300). In the present research, however, both younger and older adults appeared to differentially adjust predictions made on Trial 2. Namely, the pattern of correlations indicated that increases in between-person accuracy resulted from differentiated adjustments about the effectiveness of imagery versus repetition. For global predictions made by older adults (values for younger adults are in parentheses), correlations between Trial 1 postdictions and Trial 2 predictions made for the same strategy were .81 (.81) for imagery items and .82 (.71) for repetition items. Thus, knowledge about strategy effectiveness was used to adjust Trial 2 predictions. More important, these correlations also show a pattern of discriminant validity. Correlations within each strategy were higher than the correlations computed across the two strategies. Namely, the correlation between global postdictions for imagery items and Trial 2 predictions for repetition items was .39 (.41), and the correlation between global postdictions for repetition items and Trial 2 predictions for imagery items was .67 (.47). The pattern was the same for JOLs, but the correlations were lower. The important point is that older and younger adults adjusted Trial 2 predictions to reflect differences in the effectiveness of imagery versus repetition.

In conclusion, although knowledge updating was not necessarily reflected in the absolute accuracy of metacognitive judgments, it was evident for both age groups. These results do not support the hypothesis that older adults are deficient in using metacognitive monitoring to update knowledge about strategy effectiveness. We offered several hypotheses for why our conclusions were discrepant with those from previous research, which may guide future research in exploring how people learn about strategy effectiveness from task experience.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 88*, 238–246.

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.

Bruce, P. R., Coyne, A. C., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology, 37*, 354–357.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50–71.

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., & 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.

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

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging, 5*, 215–227.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 417–466.) Mahwah, NJ: Erlbaum.

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.

Kausler, D. H. (1994). *Learning and memory in normal aging.* New York: Academic Press.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217–273.

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.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1083–1119.

Lachman, M. E., Steinberg, E. S., & Trotter, S. D. (1987). Effects of control beliefs and attributions on memory self-assessments and performance. *Psychology and Aging, 2*, 266–271.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs, 76* (1, Pt. 2).

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah, NJ: Erlbaum.

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20*, 216–230.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Wallsten, T. S. (1996). An analysis of judgment research analyses. *Organizational Behavior and Human Decision Processes, 65*, 220–226.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Erlbaum.

Received June 18, 1999
Revision received December 7, 1999
Accepted December 16, 1999 ■