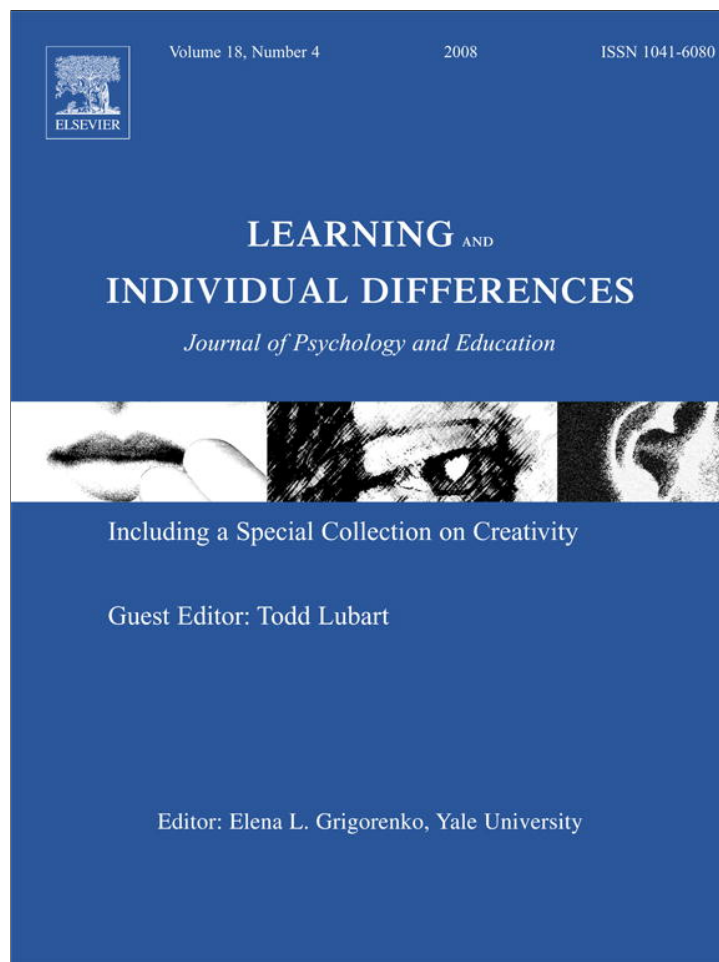


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com

Learning and Individual Differences 18 (2008) 430–445

**Learning and
Individual Differences**

www.elsevier.com/locate/lindif

How is knowledge generated about memory encoding strategy effectiveness?

Christopher Hertzog^{a,*}, Jodi Price^a, John Dunlosky^b

^a *Georgia Institute of Technology, School of Psychology, Atlanta, GA 30332-0170 USA*

^b *Kent State University, Department of Psychology, Kent, Ohio 44242-0001 USA*

Received 10 July 2007; received in revised form 1 November 2007; accepted 8 December 2007

Abstract

This study evaluated how people learn about encoding strategy effectiveness in an associative memory task. Individuals studied two lists of paired associates under instructions to use either a normatively effective strategy (interactive imagery) or a normatively ineffective strategy (rote repetition) for each pair. Questionnaire ratings of imagery effectiveness increased and ratings of repetition effectiveness decreased after task experience, demonstrating new knowledge about strategy effectiveness. Cued recall confidence judgments, measuring confidence in recall accuracy, were almost perfectly correlated with actual recall and strongly correlated with postdictions—estimates of recall for each strategy. A structural regression model revealed that postdictions mediated both changes in second-list predictions and changes in strategy effectiveness ratings, implicating accurate performance estimates based on item-level monitoring as the key to updating strategy knowledge.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Knowledge updating; Strategy; Associative learning; Metacognition; Monitoring

1. Introduction

Cognitive self-regulation involves selecting strategies appropriate to a task so as to realize one's task goals (e.g., Lemaire & Siegler, 1995; Schunn & Reder, 2001; Siegler & Stern, 1998). According to metacognitive theories, beliefs about the task, oneself, and the repertoire of strategies one has available all can influence initial strategy selection, and whether an individual continues a strategic approach or alters it in the face of performance-goal discrepancies (Bandura, 1997; Dunlosky & Hertzog, 1998b; Pintrich, Wolters, & Baxter, 2000). Hence, knowledge about various strategies and their effectiveness is an important aspect of self-regulation during learning that can influence learning (Hertzog & Dunlosky, 2004; Nelson & Narens, 1990; Schneider & Pressley, 1997; Winne, 1996).

This study examines individual differences in the degree to which people learn about strategy effectiveness, namely, use of an imagery mnemonic for learning new associations between

words. A number of strategies exist for forming new association between specific items. If words are associatively related (e.g., SALT–SUGAR), almost any technique, including repetition, will facilitate learning (Dunlosky & Hertzog, 1998a); but for unrelated words (e.g., TICK–SPOON), a relational mnemonic, such as interactive imagery, that binds the words in an integrated representation is highly effective (Bower, 1970; Paivio, 1978; Richardson, 1998). Mediational strategies, including interactive imagery or sentence generation, lead to superior cued recall when evaluated relative to using no strategy or to using a technique like rote repetition—simply repeating the words aloud. This paper focuses on the role that monitoring cued recall plays in learning about differential strategy effectiveness, and how these differences translate into updated declarative knowledge that could, in principle, be accessed to guide strategy selection at a different point in time, or in a different task context.

In the remainder of the Introduction, we describe a metacognitive framework about knowledge updating during task experience that frames the important questions addressed in the present study. This framework assumes that gaining accurate knowledge about strategies is based on accurately monitoring the effects of those strategies during study and test. We discuss

* Corresponding author. Tel.: +1 404 894 6774; fax: +1 404 894 8905.

E-mail addresses: christopher.hertzog@psych.gatech.edu (C. Hertzog), jodi.price@gatech.edu (J. Price), jdunlosk@kent.edu (J. Dunlosky).

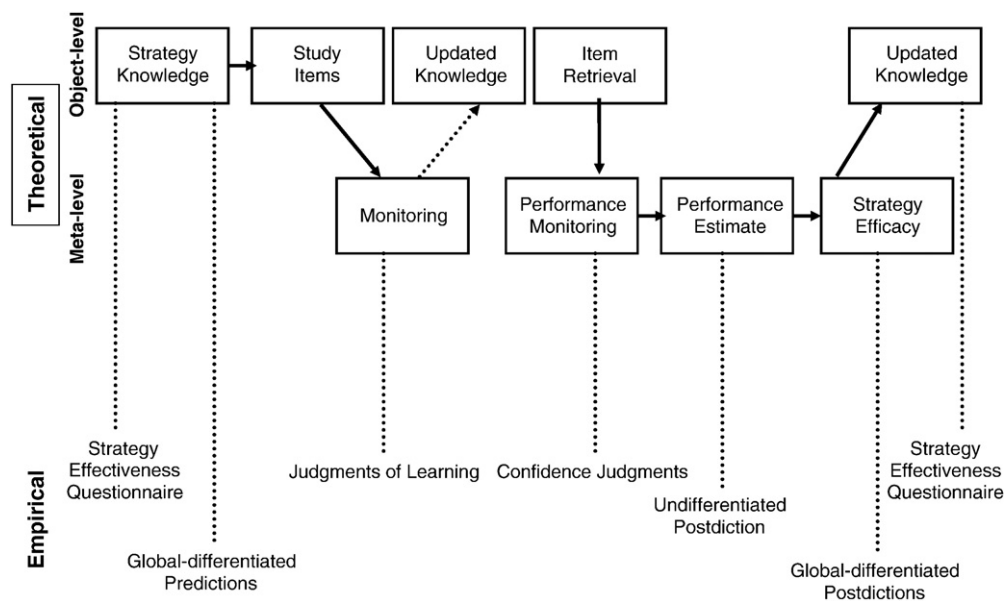


Fig. 1. Conceptual model of strategy knowledge updating. The boxes in the top section represent conceptual/theoretical processes. Dashed lines below the boxes link methods of measuring change in constructs such as knowledge updating, including different types of metacognitive judgments.

evidence that supports the framework. In doing so, our goal is to present a more complete model of acquiring declarative strategy knowledge that serves as the basis for understanding the empirical analyses of individual differences in monitoring processes and strategy knowledge updating that follow.

1.1. A metacognitive model for learning about strategy effectiveness

Fig. 1 presents a conceptual framework for how individuals learn about the effectiveness of different strategies in a paired associate (PA) learning task. Two critical assumptions of the framework are that (a) encoding strategies vary in relative effectiveness for later memory, and (b) individuals experience and evaluate variability in strategy effectiveness by monitoring their study experiences and test outcomes. More specifically, such strategy variability (Siegler & Stern, 1998) is considered an important part of creating knowledge about differential strategy effectiveness (Crowley, Shrager, & Siegler, 1997; Lovett & Schunn, 1999; Pressley, Levin, & Ghatala, 1984). Given variability, knowledge updating requires that individuals (a) monitor encoding strategies during the time of learning and/or monitor performance outcomes when tested for new associative learning, and (b) link these monitoring outcomes to the encoding strategies. Thus, monitoring of encoding and retrieval processes during study and test are central to this metacognitive framework (e.g., Nelson & Narens, 1990; Winne, 1996). A critical question is: At what point during task experience do people gain knowledge about strategy effectiveness?

The framework in Fig. 1 contains multiple pathways by which knowledge updating could occur. It also includes how different metacognitive judgments and questionnaires can be used to evaluate strategy knowledge at different points in the PA task.

1.1.1. Pre-experimental strategy knowledge

A starting point is the knowledge that people already have about strategy effectiveness before the experiment, which can be measured by questionnaire ratings of strategy effectiveness (e.g., Hertzog & Dunlosky, 2004). We argue that maximal validity will be attained by measuring knowledge about strategies that are specifically appropriate to the target task (Hadwin, Winne, Stockley, Nesbit, & Woszczyzna, 2001; Hertzog & Dunlosky, 2006). The present study uses a questionnaire designed to assess people's beliefs about the effectiveness of strategies for associative learning, which is first administered before the task begins.¹ People often show minimal pre-experimental knowledge about mediational strategies and their effectiveness (Hertzog & Dunlosky, 2006), making it possible to track *newly gained* knowledge about the strategies as people experience using them.

Prior knowledge can also influence initial *global predictions* about overall performance on the task. Metacognitive research has a long history of asking people to make global predictions. Predictions made prior to task experience (e.g., how many items will I remember on this memory test?) are notoriously inaccurate (Herrmann, Grubs, Sigmundi, & Grueneich, 1986; Hertzog, Dixon, & Hultsch, 1990; for an exception, see Ackerman, Beier,

¹ It remains possible that asking individuals to rate the effectiveness of different strategies before beginning the paired associate learning task and to provide item-level strategy reports after studying each item could have reactive effects and result in greater or different strategy use than would occur if strategies were not rated and reported. However, we intentionally altered strategy use by instructing participants to use either imagery or repetition to study each item, which provided structured experience using each strategy and allowed examination of the point at which knowledge updating occurred.

& Bowen, 2002), with individuals frequently overestimating their performance before task experience, but then exhibiting underconfidence in predictions after task exposure (i.e., the underconfidence with practice effect; Koriat, Sheffer, & Ma'ayan, 2002). Global predictions are probably influenced by multiple factors, including memory self-concept, performance goals, and personality constructs like conscientiousness and neuroticism (Hultsch, Hertzog, Dixon, & Davidson, 1988; Pearman & Storandt, 2005; West & Yassuda, 2004). For the purposes of this paper, we do not dwell on the role of beliefs, goals, and other motivational aspects that are likely to influence both performance and people's accuracy in forecasting their performance. In principle, predictions can also be influenced by an individual's knowledge about strategies that could be used on a task, as well as self-knowledge, in the form of knowing how well one has done with specific strategies in the past (e.g., Flavell, 1979).

The framework in Fig. 1 focuses on how people learn about the differential benefits of effective and ineffective strategies, but it also must take into account what people already know about strategy effectiveness. To assess the influence of strategy knowledge on performance beliefs and expectations for *specific* strategies, individuals must predict recall performance separately for each strategy they will use. Dunlosky and Hertzog (2000) found these global *differentiated* predictions were not highly accurate, as was the case for the global *undifferentiated* predictions discussed above.

1.1.2. Strategy knowledge gained during encoding

One could argue that individuals learn about strategy effectiveness as they use strategies, before the opportunity to monitor strategy effectiveness from test experience. Monitoring of encoding processes as they occur during learning can be measured by judgments of learning (JOLs) for each item, which are collected immediately after an item has been studied. JOLs should reflect any strategy knowledge that was gained during study, provided that the strategy used is a cue that is considered when making JOLs. Early metacognitive research suggested individuals were unaware that mnemonic strategies are superior to repetition because JOLs about how well items had been learned were not sensitive to the strategies that had been used (e.g., Shaughnessy, 1981; see Koriat, 1997; Lovelace, 1990). Consistent with these studies, Dunlosky and Hertzog (2000) found that JOLs were weakly influenced by the strategy used for encoding, even after encoding items on the second of two lists (see also Dunlosky & Nelson, 1994). These outcomes may speak more to the sensitivity of JOLs to strategy knowledge than to the existence of that knowledge.

One can also measure updated strategy knowledge after study with a second set of global differentiated predictions, this time made after study (but before the recall test). Poststudy predictions have increased correlations with subsequent recall (e.g., Connor, Dunlosky, & Hertzog, 1997; Hertzog, Saylor, Fleece, & Dixon, 1994; but see Lineweaver & Hertzog, 1998), relative to the correlations generated from prestudy predictions, indicating poststudy predictions may be more sensitive to encoding strategy differences.

1.1.3. Metacognitive processes during and after test performance

Learning about strategy effectiveness appears to involve monitoring performance outcomes and associating success or failure to the processing strategies that have been employed (e.g., Hunter-Blanks, Ghatala, Pressley, & Levin, 1990; Pressley et al., 1984).² Performance monitoring during recall tests is often highly accurate—individuals are able to discriminate successes from failures and then accurately estimate the relative probability of success (Dunlosky & Hertzog, 2000; Higham, 2002; Matvey, Dunlosky, Shaw, Parks, & Hertzog, 2002). Accurate performance monitoring improves individuals' accuracy in predicting levels of future performance. In multiple-trial experiments with a single list, prediction accuracy improves substantially after the first study-test trial (e.g., Finn & Metcalfe, 2007; Koriat, 1997; Koriat & Bjork, 2006). Koriat and Bjork (2006) demonstrated that experiencing multiple tests, rather than multiple study opportunities, was the biggest influence on improvements in the accuracy of item-level JOLs. In Koriat and Bjork's terms, individuals shift the basis of their metacognitive judgments from theories about how well they will perform on a given item to experience with how they actually have performed on it. Performance monitoring, in this sense, is the engine that drives this process of discovery.

Because acquiring new strategy knowledge requires actual memory test experience to demonstrate the differential benefits of various strategies (e.g., Hunter-Blanks et al., 1990), performance monitoring is likely to be a critical influence on strategy knowledge updating. Testing this hypothesis requires measurement of performance monitoring and inferences about strategy effectiveness based on that monitoring. On-line performance monitoring at the time of item recall can be measured by confidence judgments (CJs) taken after each recall attempt. Dunlosky and Hertzog (2000) found that individuals' CJs were highly correlated with PA item recall within individuals (i.e., showed high intraindividual Goodman–Kruskal gamma correlations; Nelson, 1984), indicating near-perfect discrimination of recall accuracy at the level of individual PA items (see also Higham, 2002). Item-level performance monitoring serves as the potential basis for inferences about how many items were recalled with each strategy.

However, accurate performance monitoring is not sufficient for obtaining new strategy knowledge. Individuals must also attribute successful remembering to the strategies used for encoding the items (Devolder & Pressley, 1992; Fabricius & Cavalier, 1989; Fabricius & Hagen, 1984; Schneider & Pressley, 1997), associate specific recall outcomes with prior strategy use,

² According to Crowley et al. (1997), strategy learning can be considered to be relatively mechanistic and automatic, perhaps occurring without conscious awareness (Siegler & Stern, 1998). They contrasted associative learning mechanisms with what they termed metacognitive learning processes, which they characterized as being intentional, reflective, and requiring analysis of both task situation and the strategic repertoire to enable adaptive strategic choices (see also Kuhn, 2001). This distinction appears to be useful, but we claim that incidental associative learning mechanisms are an insufficient basis for generating accurate declarative knowledge about differential strategy effectiveness after intentional learning instructions in a PA task.

and aggregate this information across multiple items to produce declarative knowledge about the base-rate of strategy effectiveness (Winne, 1996).

The framework in Fig. 1 suggests that the first phase of the process linking performance monitoring to strategy knowledge could be an assessment of overall level of performance. We shall refer to it as *performance estimation*. It can be measured by postdictions collected after the conclusion of the recall test, in which the investigator asks individuals to estimate how many items (or percentage of items) were remembered. Postdictions are often highly accurate when made across the entire list (e.g., Devolder, Brigham, & Pressley, 1990; Hertzog et al., 1994; Pressley & Ghatala, 1989). Postdictions and CJs are not merely alternative methods of measuring performance monitoring. Results from Dunlosky and Hertzog (2000) argue for a meaningful differentiation of these two post-performance judgments, and probably, the underlying processes that generate them. Postdictions over an entire set of items, such as a list of words, require inferential processes to estimate the number (or percentage) of items successfully recalled.

One could generate accurate postdictions through an exhaustive search of memory for instances of memory success or failure, assuming that all performance outcomes were accessible at the end of the test. However, when providing a global postdiction individuals may not attempt to engage in extensive retrieval-based sampling of item strategies or outcomes before estimating the probabilities (Winne, 1996). To the extent that instance retrieval is used, it may be limited in scope and biased by an availability or representativeness heuristic (e.g., Brown, 1995; Kahneman, 2003; Manis, Shedler, Jonides, & Nelson, 1993). Moreover, individuals may not attempt to retrieve any specific instances of success with each strategy. Instead, individuals appear to have rapid access to an intuitive, non-analytic, fallible (but largely accurate) sense of how many *total* items they had successfully recalled, irrespective of how those items had been processed at encoding and retrieval. Some have claimed that event frequencies are stored automatically and accurately (e.g., Hasher & Zacks, 1984), resulting in well-calibrated memory for relative degree of success (see also White & Wixted, 1999). Such information could be the basis for generating accurate postdictions from on-line performance monitoring.

Thus, our theoretical scheme treats postdictions as inferential in nature and separable from concurrent performance monitoring. In Fig. 1, we treat postdictions as measures of a first stage inference about monitoring, performance estimation, that depends directly on item-level performance monitoring (as reflected in CJs) but is also influenced by other processes.

Performance estimates could be highly accurate, yet inferences about contributions of different strategies to this overall level of performance could be highly inaccurate. To acquire declarative strategy knowledge, individuals must associate test outcomes with encoding strategies to infer the degree of benefit for each strategy, or of one strategy relative to another. We refer to this concept as *subjective strategy efficacy* (see Fig. 1).

Global differentiated postdictions measure in part the outcome of translating performance monitoring into evaluations

of differential strategy effectiveness (Dunlosky & Hertzog, 2000). They differ from the typical postdiction collected in prior research, *undifferentiated postdictions*, which ask for estimates of recall over the entire list, without reference to subtypes of items (whether classified by strategy or any other dimension of item and/or study characteristics).

If information from performance monitoring included complete information about strategies used during encoding, then one could expect subjective strategy efficacy to be equally accurate (i.e., differentiated postdictions would be as accurate as the undifferentiated postdiction). However, it appears likely that accuracy is lost when the differentiated estimates are made (Dunlosky & Hertzog, 2000)—suggesting a distinction between performance estimation and estimated strategy efficacy.

Why would this be so? Access to prior encoding strategies at the time of test is likely to be fallible and subject to interference and distortion (Winne, 1996). Given that information must be aggregated over multiple items studied with different strategies, inferences under uncertainty about prior item encoding history probably play a role in abstracting strategy knowledge from performance monitoring outcomes (Maley, Hunt, & Parr, 2000). Furthermore, use of heuristic inferential processes to judge recall outcomes with different strategies could also introduce distortion (Brown, 1995; Cavanaugh, Feldman, & Hertzog, 1998; Gigerenzer, Todd, & the ABC Research Group, 1999).

The framework in Fig. 1 stipulates that accurate performance monitoring, as measured by CJs, influences performance estimates, as measured by an undifferentiated postdiction. In turn, performance estimation influences the accuracy of abstracted information about the base-rate of success with each strategy. This abstracted evaluation of encoding strategy effectiveness then mediates any increases in declarative knowledge about normative strategy effectiveness.

1.1.4. Metacognitive judgments on a second list

Instead of multiple study-test trials with the same list, in which item learning can be studied along with metacognitive judgments (e.g., Finn & Metcalfe, 2007; Koriat, 1997; Koriat & Bjork, 2006), we used a single study-test trial with two different but equivalent item lists. The use of different lists allowed us to measure changes in metacognitive ratings (i.e., differentiated predictions or JOLs) and declarative strategy knowledge (i.e., knowledge updating) as a function of task experience.³ With respect to our framework in Fig. 1, changes in between-person correlations of judgments with strategies, or of within-strategy judgments with recall, could indicate increased knowledge of strategy effects on memory.

1.1.4.1. Trial 2 predictions. The accuracy of global predictions, reflected in correlations of predictions with list recall,

³ Changes in correlations of judgments with recall reflect strategy knowledge influences and other factors, such as proactive interference from List 1 to List 2 at the time of List 2 test. We assume such effects to be nil in our analyses. Regarding interference effects, our lists are structured to be devoid of between-cue associations (i.e., forward and backward association strength across all lists was on average zero) to minimize any such effects.

improves across multiple lists as a function of task experience (Herrmann et al., 1986; Hertzog et al., 1990, 1994). This improvement may be understood as a consequence of accurate performance estimation, because undifferentiated postdictions correlate strongly with performance and with Trial 2 predictions (Connor et al., 1997; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Hertzog et al., 1994). Such an improvement may say little about changes in declarative strategy knowledge per se.

However, the strategy-differentiated predictions completed before studying List 2 should indirectly reflect strategy knowledge updating. If so, then Trial 2 predictions for each strategy should be more accurate than Trial 1 predictions collected prior to task experience. Dunlosky and Hertzog (2000) found such effects; Trial 2 correlations of differentiated predictions were higher than Trial 1 correlations. However, one should not presume that new, durable strategy knowledge has been extracted unless this correlational difference is greater than what could be expected based on an improvement in global performance estimation as measured by an undifferentiated postdiction.

1.1.4.2. Trial 2 JOLs. To the extent that strategy knowledge influences JOLs, then one would predict greater differentiation of JOLs for effective strategies than ineffective strategies at Trial 2. Dunlosky and Hertzog (2000) found some evidence for increased influence of strategies used on Trial 2 JOLs, but the degree of upgrading was far from impressive. For example, unlike differentiated predictions, as compared to Trial 1 recall, Trial 2 mean JOLs showed a much reduced tendency for higher correlations with Trial 2 recall. Such findings suggest that JOLs are not highly influenced by the strategy used, even after test experience, which has actually been suggested in earlier literature (Koriat, 1997; Lovelace, 1990). It is therefore unfortunate that studies have relied upon JOLs as the primary measures of strategy knowledge updating (e.g., Bieman-Copland & Charness, 1994; Shaw & Craik, 1989).

1.1.4.3. Post-task strategy knowledge questionnaire. Repeating the strategy knowledge questionnaire provides information about whether people have learned about the differential effectiveness in a manner that is likely to survive the experimental session. As noted already, JOLs may not be sensitive to updating, and even global differentiated postdictions may not entirely reflect the magnitude of the strategy effects (Dunlosky & Hertzog, 2000). Nevertheless, a metacognitive view of strategy knowledge updating hypothesizes that subjective strategy effectiveness, as measured by the differentiated postdictions, will have a strong effect on the questionnaire ratings, controlling on pre-experimental questionnaire ratings. If so, this would be strong evidence that declarative knowledge developed through task experience. Moreover, because some metacognitive judgments may be insensitive to strategy knowledge updating, reassessment of questionnaire ratings of strategy effectiveness provides a critical means of determining whether knowledge updating occurs within a session, even if it is not reflected in metacognitive judgments themselves.

2. Goals of the present study

The data we analyze in this study have been partly used in a companion paper (Hertzog et al., *in press*). That paper focused on the issue of absolute accuracy of global predictions and postdictions; specifically, postdictions and predictions for the imagery strategy underestimated actual recall with the imagery strategy by more than 20%. This underestimation for imagery items was repaired to a degree by an experimental manipulation—blocking tested items by reported strategy (either imagery or repetition). In terms of the framework of Fig. 1, such effects suggest that the inferential processes required to produce perceived strategy efficacy are a locus of distortion.

The present study was designed to evaluate central implications of the framework in Fig. 1 using individual differences information generated in a knowledge updating experiment. We hypothesized that accurate performance monitoring (in CJs) is associated with relatively accurate performance estimation (i.e., undifferentiated postdictions). However, converting accurate overall recall estimates into differentiated postdictions uses a fallible scaling heuristic in which estimated total recall serves as an anchor (Kahneman, Slovic, & Tversky, 1982; Scheck, Meeter, & Nelson, 2004). Individuals may access this anchor and then use their non-analytic (and probably ordinal) knowledge about the superiority of imagery to adjust their differentiated postdictions up for imagery and down for repetition (see also Cavanaugh et al., 1998). To gain preliminary evidence on this issue, we included an overall (undifferentiated) postdiction in our experiments and evaluated its relationship to the strategy-differentiated postdictions.

If general performance monitoring is the only mechanism operating, then the undifferentiated postdiction would predict the differentiated postdictions and mediate the effects of recall for each strategy on postdictions for that strategy. An alternative hypothesis was that strategy-specific performance monitoring would translate into differences in subjective strategy efficacy (as in Fig. 1) between the strategies. Under the hypothesis of strategy-specific effects, recall for each strategy should predict differentiated postdictions independent of the undifferentiated postdiction—providing evidence of subjective strategy efficacy beyond what could be expected from a general performance monitoring that is agnostic to strategy type. We also hypothesized that subjective strategy efficacy, as measured by the differentiated postdictions, would mediate changes in the questionnaire ratings of strategy effectiveness. These hypotheses, if confirmed, would support the role of performance monitoring, performance estimation, and subjective strategy efficacy in producing declarative knowledge about strategy effectiveness.

2.1. Method

The methods for this experiment are described in detail in Hertzog et al., (*in press*). We cover some of the key elements here.

2.1.1. Design

The design was a $2 \times 2 \times 3$ mixed factorial, with Trial (first vs. second study list), Strategy instructions (repetition vs. imagery),

and Condition (no undifferentiated postdiction, undifferentiated postdiction with no performance feedback, and undifferentiated postdiction, followed by performance feedback). Trial and Strategy instructions were within-subjects factors. Because the Condition factor had no impact on the data, our analysis ignores it.

2.1.2. Participants

Two hundred fifty-three undergraduate students at the Georgia Institute of Technology who were enrolled in psychology courses participated in exchange for course credit.

2.1.3. Strategy Effectiveness Questionnaire

A brief background information form collected demographic information from participants. The Personal Encoding Preferences (PEP) questionnaire measured participants' subjective ratings of the effectiveness of various memory strategies (Hertzog & Dunlosky, 2004). The PEP questionnaires list memory strategies, along with definitions and examples of how they would be used. Respondents rate the effectiveness of each strategy on a 10-point (1–10; 1=ineffective, 10=highly effective) Likert scale. Participants are also given the chance to add and rate strategies not in the list.

The PEP I was administered prior to the PA task. The PEP II assesses participants' strategy knowledge after task experience. The two versions are identical except the PEP II directs participants to imagine themselves advising a friend about the effectiveness of each of the strategies to avoid respondent reactions to a second administration of the identical questionnaire.

2.1.4. Procedure

Each of the experimental conditions consisted of two lists of 50 PA items which participants studied and on which they were tested. For each list, participants were instructed to use one of the two strategies (repetition or interactive imagery) for half the items at random. Before, during, and after studying each list, and after the recall test, individuals made the different metacognitive judgments highlighted in Fig. 1 and the Introduction.

2.1.4.1. Prestudy global differentiated predictions. Participants made separate predictions, estimating the percentage of items they would correctly recall, for imagery and repetition. The order of making these predictions (i.e., repetition first or imagery first) was counterbalanced across participants. Due to a programming error, the instruction screen for the repetition prestudy predictions in the undifferentiated postdiction with feedback condition erroneously labeled the predictions as imagery. For these 86 participants, the prestudy repetition prediction was treated as missing data.

2.1.4.2. Study. After making their predictions, participants studied each word pair (presented in a random order) for 6 s and immediately made a JOL for that item (scaled as 0–100% confidence they could recall the target word in a later test). Participants then reported their strategies by entering a number corresponding to: (1) repetition, (2) imagery, (3) a combination of both imagery and repetition, (4) a strategy other than

imagery or repetition, (5) no strategy, or (6) did not have time to use a strategy. Immediately after study, participants made a second set of differentiated predictions for the two strategies we shall refer to as poststudy differentiated predictions.

2.1.4.3. Test. At test, participants were presented with the cue words from the previously studied PA items, and typed their chosen response. Responses were scored as correct if the first three letters of the typed response matched the first three letters of the correct target word.

Following the cued recall response, participants made a CJ for that response, (scaled as 0–100% confidence in its correctness). No accuracy feedback was provided.

2.1.4.4. Global undifferentiated postdictions. Participants in the control condition did not make undifferentiated postdictions. Participants in the other two conditions estimated the total percentage of the 50 items they had remembered correctly. Participants in the feedback condition were then told the percentage of items they had correctly recalled during the test phase of the trial.

2.1.4.5. Global differentiated postdictions. All participants then estimated the percentage of PA items they had correctly recalled (out of 25) using imagery and repetition.

2.1.4.6. Trial 2. Following the differentiated postdictions, Trial 2 began with a new list of 50 PA items, following the same procedure. At the end of the differentiated postdictions for Trial 2, the PEP II questionnaire was administered.

2.2. Results

2.2.1. Compliance with strategy instructions

Students reported on average 83% (SD=15.2) and 84% (SD=17.4) compliance with imagery instructions on List 1 and List 2, respectively (see Hertzog et al., in press for more details). Across lists, reported repetition and imagery use yielded test–retest correlations of .77 and .70, respectively. This high stability in strategic behavior suggested that people's strategy use reflected, in large part, individual differences in preferences or abilities to use the two strategies. Moreover, noncompliance was selective, with less compliance on List 2 for 121212repetition instructions. A regression analysis showed that, controlling on List 1 imagery use, List 2 imagery use was reliably predicted by the after-task imagery effectiveness ratings, $\beta = .22$, $t = 5.05$, $p < .001$. Conversely, List 2 repetition use was not reliably predicted by after-task repetition effectiveness ratings, controlling on List 1 repetition use, $\beta = .03$, $t < 1$. Thus, students switched to imagery on the second list, but differentially so if they rated imagery as an effective strategy. Apparently, individuals learned that repetition was the less effective strategy and were slightly less inclined to use it. For this reason, our subsequent analyses use reported strategy (instead of instructed strategy) to classify relevant variables.

Table 1
Means, standard deviations, and correlations from Experiment 1

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	
1. PEP-I Imagery Rating	1													
2. PEP-I Rote Rating	-.24**	1												
3. List 1 Imagery Prediction	.21**	-.13*	1											
4. List 1 Rote Prediction	-.12	.17*	.65**	1										
5. List 1 Imagery JOL	.22**	.03	.57**	.47**	1									
6. List 1 Rote JOL	.02	.15*	.45**	.58**	.81**	1								
7. List 1 Imagery Poststudy	.25**	-.07	.62**	.33**	.67**	.57**	1							
8. List 1 Rote Poststudy	.02	.10	.48**	.59**	.60**	.73**	.72**	1						
9. List 1 Imagery Recall	.19**	-.01	.26**	.11	.19**	.05	.29**	.12	1					
10. List 1 Rote Recall	.07	.10	.17**	.25**	.13*	.18**	.20**	.30**	.54**	1				
11. List 1 Imagery CJ	.23**	-.01	.31**	.16*	.29**	.13**	.39**	.21**	.94**	.54**	1			
12. List 1 Rote CJ	.10	.09	.19**	.26**	.19**	.23**	.26**	.35**	.45**	.92**	.49**	1		
13. List 1 Undiff. Postdiction	.10	-.03	.37**	.23*	.29**	.20**	.48**	.37**	.71**	.70**	.73**	.65**	1	
14. List 1 Imagery Postdiction	.18**	-.06	.36**	.19*	.27**	.12	.48**	.25**	.74**	.49**	.76**	.45**	.80**	1
15. List 1 Rote Postdiction	.11	.09	.30**	.39**	.25**	.27**	.38**	.49**	.49**	.71**	.50**	.69**	.74**	
16. List 2 Imagery Prediction	.21**	-.06	.44**	.23**	.37**	.22**	.55**	.30**	.66**	.41**	.71**	.39**	.74**	
17. List 2 Rote Prediction	.16*	.09	.29**	.40**	.34**	.39**	.38**	.59**	.35**	.64**	.42**	.68**	.69**	
18. List 2 Imagery JOL	.22**	-.00	.47**	.31**	.74**	.53**	.63**	.50**	.45**	.35**	.54**	.36**	.57**	
19. List 2 Rote JOL	.06	.13*	.36**	.42**	.57**	.66**	.52**	.64**	.33**	.49**	.38**	.51**	.50**	
20. List 2 Imagery Poststudy	.17**	-.01	.37**	.22**	.39**	.25**	.56**	.37**	.60**	.39**	.65**	.40**	.67**	
21. List 2 Rote Poststudy	.05	.11	.28**	.38**	.35**	.39**	.44**	.59**	.39**	.57**	.44**	.58**	.64**	
22. List 2 Imagery Recall	.18**	.06	.22**	.10	.16*	.05	.26**	.10	.80**	.49**	.81**	.43**	.60**	
23. List 2 Rote Recall	.03	.12	.16**	.22**	.14*	.20**	.20**	.28**	.51**	.70**	.50**	.68**	.58**	
24. List 2 Imagery CJ	.23**	.03	.26**	.12	.24**	.10	.33**	.17**	.74**	.47**	.81**	.46**	.60**	
25. List 2 Rote CJ	.04	.15*	.17**	.28**	.18**	.22**	.25**	.32**	.44**	.65**	.49**	.68**	.57**	
26. List 2 Undiff. Postdiction	.09	.03	.36**	.33**	.30**	.26**	.43**	.40**	.64**	.63**	.67**	.64**	.76**	
27. List 2 Imagery Postdiction	.18**	-.03	.33**	.18*	.25**	.13*	.40**	.25**	.68**	.48**	.71**	.45**	.73**	
28. List 2 Rote Postdiction	.07	.09	.28**	.35**	.23**	.26**	.32**	.46**	.46**	.66**	.49**	.64**	.67**	
29. PEP-II Imagery Rating	.38**	-.01	.17**	-.09	.21**	.03	.25**	.06	.41**	.14*	.47**	.14*	.32**	
30. PEP-II Rote Rating	-.22**	.37**	-.04	.26**	.02	.14*	-.01	.25**	.05	.36	.04	.34**	.20**	
Means	6.88	58.05	46.32	50.83	37.01	46.55	31.81	.68	.30	70.05	33.11	44.96	55.15	
SD	2.05	1.94	22.41	19.77	18.38	20.61	18.21	.22	.22	20.24	21.44	24.49	26.98	

Abbreviations: Imagery=interactive imagery; Rote=rote repetition; Prediction=global prestudy prediction; JOLs=judgments of learning; Poststudy=global poststudy prediction; CJ=confidence judgments; Undiff.=Undifferentiated; Postdiction=global differentiated postdiction; SD=standard deviation.

**= $p < .001$.

*= $p < .05$.

2.2.2. Individual differences in strategy ratings, recall, and metacognition

Table 1 reports the means, standard deviations, and correlations for the PEP strategy effectiveness ratings, metacognitive judgments, and PA recall for both lists, using pairwise deletion of missing data. Large differences occurred in PA recall for items reported as studied with imagery versus repetition for both lists. Scaled as Cohen's (1988) d statistic, differences in standard deviation units were $d=1.71$ for List 1 and $d=1.67$ for List 2. Cohen argued that $d > .8$ represents a large effect size. Clearly, the strategies differed substantially in their effectiveness.

The questionnaire data indicated individuals learned about this differential strategy effectiveness. Prior to task experience,

the rated effectiveness of the two strategies was quite similar. After task experience, imagery was rated as far superior to repetition, $d=2.23$, reflecting greater strategy knowledge. The focus of our paper is in understanding how this updating in strategy knowledge occurred.

2.2.3. Global differentiated predictions and JOLs

Fig. 2 highlights the increases in Pearson correlations of the magnitude of individuals' predictions and mean JOLs with PA recall across lists with selected correlations from Table 1. Learning about strategy knowledge during List 1 influenced the alignment of individual differences in predictions with recall. The largest increase in correlations occurred between Trial 1 and Trial 2 (i.e., after recall experience). In particular, prestudy

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
.58**	1															
.88**	.48**	1														
.49**	.88**	.53**	1													
.58**	.46**	.66**	.51**	1												
.35**	.61**	.45**	.73**	.71**	1											
.78**	.53**	.81**	.52**	.69**	.50**	1										
.50**	.81**	.49**	.84**	.56**	.72**	.63**	1									
.65**	.45**	.63**	.36**	.43**	.31**	.61**	.41**	1								
.39**	.58**	.33**	.52**	.27**	.45**	.34**	.50**	.51**	1							
.65**	.44**	.66**	.40**	.50**	.35**	.64**	.45**	.94**	.49**	1						
.38**	.57**	.35**	.56**	.31**	.50**	.37**	.55**	.49**	.42**	.53**	1					
.70**	.70**	.69**	.67**	.55**	.54**	.70**	.67**	.69**	.67**	.71**	.68**	1				
.81**	.55**	.78**	.49**	.57**	.40**	.77**	.54**	.73**	.51**	.73**	.53**	.82**	1			
.52**	.81**	.46**	.75**	.45**	.58**	.51**	.81**	.49**	.64**	.51**	.66**	.75**	.65**	1		
.44**	.18**	.43**	.17**	.34**	.16*	.34**	.16*	.47**	.17**	.49**	.20**	.32**	.46**	.21**	1	
-.01	.39**	-.05	.38**	.06	.32**	-.02	.39**	.08	.30**	.06	.30**	.28**	-.02	.45**	-.13*	1
29.45	53.55	30.50	46.54	28.36	45.07	26.58	.69	.32	70.91	33.98	45.54	51.51	29.58	8.67	4.64	
23.64	24.36	21.26	21.24	18.62	23.69	19.52	.22	.22	21.55	22.48	23.69	26.91	23.71	1.24	2.24	

predictions and poststudy predictions correlated about the same with later recall for both lists. By contrast, poststudy predictions for List 1 were not as highly correlated with recall as the List 1 postdictions. This pattern suggested the bulk of knowledge updating from List 1 to List 2 reflected mechanisms operating after List 1 test experience instead of during item study, prior to the test.

One way to capture change in the metacognitive judgments is to evaluate the stability of individual differences in the judgments from List 1 to List 2. High correlations across the trials would suggest little change in the judgments. The List 1 and List 2 differentiated prestudy predictions for imagery correlated .44 for imagery and .40 for repetition, suggesting individuals realigned their predictions after task experience. In

contrast, the stability of JOLs across the two lists was much higher—the correlations were .74 for imagery and .66 for reported repetition items. This outcome suggested that individual differences in the global predictions shifted more dramatically as a function of test experience than was the case for JOLs.

2.2.4. Performance monitoring and inferential processes after test

The correlations among the postdictions, CJs, and recall suggested that performance monitoring was relatively accurate in aligning individual differences for judgments and recall. Broken down by strategy report, the correlations of mean CJs and recall were all above .90. The undifferentiated postdictions correlated .82 and .83 with List 1 and List 2 total recall,

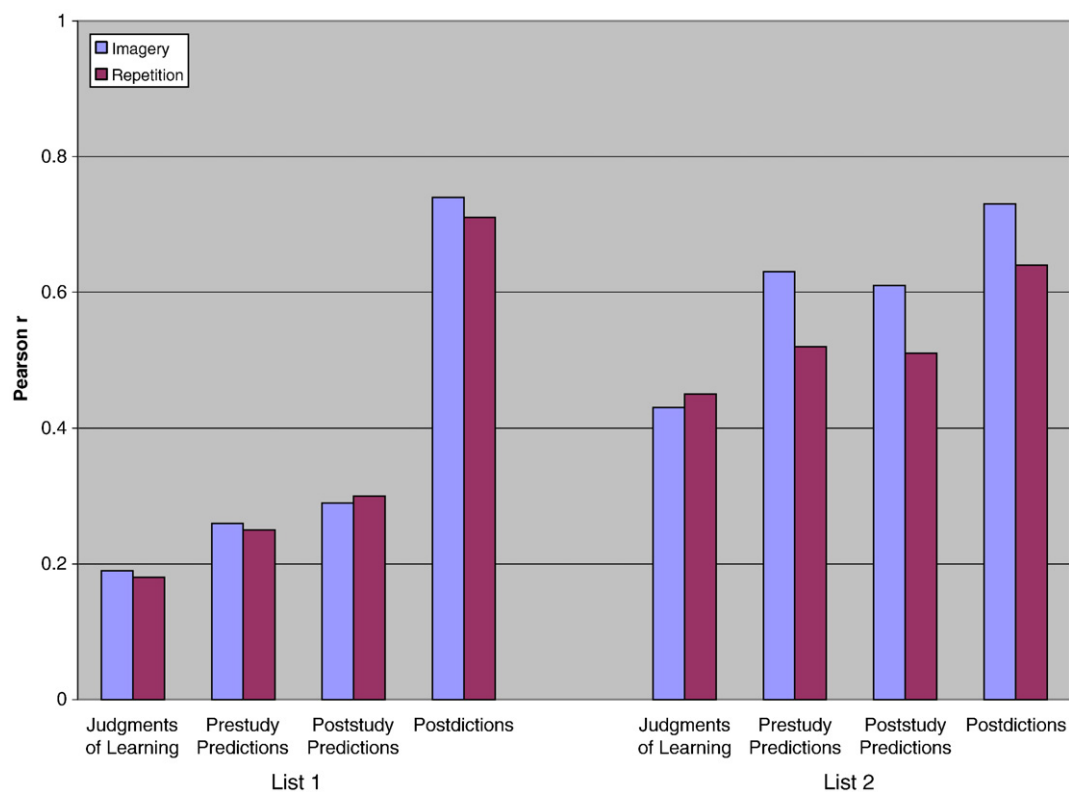


Fig. 2. Pearson r correlations calculated between (a) individuals' mean judgments of learning and mean recall performance, (b) individuals' global differentiated prestudy predictions and mean recall performance (c) individuals' global differentiated poststudy predictions and mean recall performance, and (d) individuals' global differentiated postdictions and mean recall performance. For each list, correlations were calculated separately for imagery and rote repetition strategy use.

respectively, indicating some drop-off in correlations from mean CJs. The correlations of differentiated postdictions and recall were slightly lower than that for the undifferentiated postdictions with recall (see Table 1). This was not merely a function of lower reliability in the strategy-segregated recall measures. The relationships of differentiated postdictions with recall remained lower when they were correlated with total recall (List 1 correlations of .72 and .69 for imagery and repetition postdictions, respectively; List 2 correlations of .76 and .67 for imagery and repetition postdictions, respectively). Hence the lower correlations, relative to the undifferentiated postdictions, were not merely a byproduct of differential test length and reduced reliability when dividing recall into subsets based on encoding strategy.

None of these correlations varied materially between the lists, suggesting that the realignment of individual differences occurred after the recall test for List 1 and was simply maintained during List 2.

2.2.5. Structural regression model

We conducted a structural regression model to evaluate the shifts in correlations within and between lists. The central premise of the model is that knowledge updating is reflected in changes in List 2 measures of knowledge (e.g., predictions and PEP scores) that are mediated by accurate performance monitoring, as reflected in List 1 postdictions. The model used the PEP ratings for imagery and repetition, differentiated predictions for both lists, the undifferentiated postdiction for

List 1, differentiated postdictions for List 1, and List 1 recall as variables. Dunlosky and Hertzog (2000) previously showed that one can model CJs as mediating the effects of recall on postdictions. Given the near-perfect correlations of CJs with recall, we deemed it unnecessary to replicate this mediation effect in the present study, and omitted the CJ variables in order to leave the model as uncomplicated as possible.⁴

The model was used to evaluate several hypotheses. First, we expected that overall levels of PA recall would influence the undifferentiated postdiction (consistent with the high correlations just reported, and earlier literature). We further expected that the undifferentiated postdiction would have strong paths to the differentiated postdictions. Most important, the model tested the hypothesis of accurate performance monitoring for each strategy, above and beyond the accuracy of the undifferentiated postdiction, as noted above. If so, performance monitoring specific to each strategy would be reflected in reliable paths from imagery and repetition recall to imagery and repetition postdictions, controlling for (a) the relation of overall recall to the undifferentiated postdiction and (b) the relation of the undifferentiated postdiction to the postdictions for performance with each strategy. In contrast, a hypothesis that performance

⁴ We also opted not to report model results with mean JOLs instead of global predictions. Dunlosky and Hertzog (2000) showed that mean JOLs are more stable over lists, and show lower evidence of postdictions influencing List 2 JOLs, relative to the global predictions. The correlations reported above replicate this pattern.

monitoring consists solely of awareness of overall levels of performance, such as through an intuitive sense of frequency of overall recall outcomes, would argue that the undifferentiated postdiction would completely mediate the effect of strategy-specific recall on strategy-differentiated postdictions.

To evaluate this hypothesis, the structural regression model specified a latent recall variable that was defined by the unit-weighted sum of the two strategy-specific recall variables. This latent total recall variable was modeled as the proximal determinant of the undifferentiated postdiction. This kind of specification (analogous to a MIMIC model in SEM; see Jöreskog & Sörbom, 2003) enabled us to account for the general recall effect on the undifferentiated postdiction, while maintaining the strategy-specific recall variables as determinants of differentiated postdictions.

Second, we hypothesized that pre-experimental knowledge (or beliefs) about the effectiveness of the two strategies, as measured by PEP-I ratings, would influence the first predictions for the imagery and repetition strategies. More important, we predicted that the accuracy of performance monitoring would shape subjective strategy efficacy and thereby determine the degree of knowledge updating manifested at the end of the experiment. Hence we hypothesized that the changes in effectiveness ratings after task experience would be predicted by the differentiated postdictions (which would, in turn, mediate effects of recall on the effectiveness ratings). This hypothesis implicates performance monitoring and subsequent inferential mechanisms as the major influence on strategy knowledge updating. It can be contrasted against a hypothesis that recall performance itself is the better predictor of changes in effectiveness ratings that is carried by some other, unmeasured mechanism.

We used full-information maximum likelihood estimation in LISREL 8.53 (Jöreskog & Sörbom, 2003) to estimate the parameters and associated standard errors. The program also provided a likelihood-ratio χ^2 test of model fit to the data, along with two goodness-of-fit indices, the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), and the Comparative Fit Index (CFI; Bentler, 1990). RMSEAs of .05 or smaller, and CFIs of .95 or greater are considered indicative of an excellent degree of fit of the model to the data.

The initial model we specified using this logic fit the data relatively well, $\chi^2=74.17$, $df=53$, $p<.03$, RMSEA=.067, CFI=.977. However, there were some indications of misspecification, reflected in large standardized residuals and large modification indices for some fixed parameters. Based on these results we added a few regression parameters to the model, and trimmed a few nonsignificant paths. The revised model provided an excellent fit to the data, $\chi^2=64.94$, $df=54$, $p<.15$, RMSEA=.042, CFI=.988. Table 2 reports unstandardized regression coefficients and their standard errors; Fig. 3 graphs the standardized regression coefficients (direct effects).

The first major finding was that overall level of recall had a strong relationship to the undifferentiated postdiction, which in turn had strong paths to both differentiated postdictions. This suggested that monitoring overall levels of recall played a key role in determining the strategy-differentiated postdictions. As

Table 2

Unstandardized regression coefficients (with standard errors) from accepted structural equation model Experiment 1

Path	Estimate (SE)
Imagery prediction List 1 → Imagery recall List 1	2.84 (.27)
Rote prediction List 1 → Rote recall List 1	1.94 (.18)
Imagery recall List 1 → Imagery postdiction List 1	3.27 (.28)
Rote recall List 1 → Rote postdiction List 1	3.26 (.35)
Imagery recall List 1 → Rote postdiction List 1	-2.71 (-.25)
Rote prediction List 1 → Rote postdiction List 1	3.90 (.30)
Imagery postdiction List 1 → Imagery prediction List 2	19.54 (.77)
Imagery prediction List 1 → Imagery prediction List 2	4.04 (.19)
Rote postdiction List 1 → Rote prediction List 2	18.21 (.79)
Rote prediction List 1 → Rote prediction List 2	-.07 (-.01)
Rote postdiction List 1 → Rote rating on post-task PEP	3.70 (.03)
Rote postdiction List 1 → Imagery rating on post-task PEP	-2.39 (-.01)
Imagery postdiction List 1 → Imagery rating on post-task PEP	4.94 (.02)
Imagery rating on pre-task PEP → Imagery prediction on List 1	3.89 (3.20)
Rote rating on pre-task PEP → Rote prediction List 1	2.96 (2.16)
Undifferentiated postdiction List 1 → Imagery postdiction List 1	9.64 (.80)
Undifferentiated postdiction List 1 → Rote postdiction List 1	6.46 (.62)
Recall List 1 → Undifferentiated postdiction List 1	10.15 (.49)
Rote rating on pre-task PEP → Rote rating on post-task PEP	2.65 (.25)
Imagery rating on pre-task PEP → Imagery rating on post-task PEP	3.00 (.15)

important, the second (and more unique) major finding was that there were also reliable paths from imagery recall to the imagery postdiction and from repetition recall to the repetition postdiction, suggesting that monitoring strategy specific effects also influenced postdictions. Indeed, an alternative model that *deleted* the effects of strategy-specific recall on differentiated postdictions had a poorer fit to the data, $\chi^2=89.78$, $df=57$, $p<.004$, RMSEA=.077, CFI=.965. Thus, accurate performance monitoring also apparently occurred at the level of specific strategies and influenced subjective strategy efficacy.

The third major finding was the low stability in predictions from List 1 to List 2, with essentially no stability in repetition predictions across the two lists. Conversely, there were strong paths from the differentiated postdictions to the List 2 predictions for each strategy, controlling on the List 1 predictions. These findings demonstrate substantial knowledge updating, with more accurate predictions for List 2 (as expected from the differences in correlations shown in Fig. 2). Moreover, this pattern showed that this knowledge updating was fully mediated by the differentiated postdictions.

Finally, a similar updating effect was observed for the strategy effectiveness ratings. The stability of these ratings from pre-task to post-task was low, with strong paths from the differentiated postdictions to the post-task effectiveness ratings. An alternative model which specified these effects to be driven by imagery recall and repetition recall, while eliminating the paths from the postdictions to the effectiveness ratings, also resulted in poorer fit, $\chi^2=78.40$, $df=55$, $p<.03$, RMSEA=.069, CFI=.975. This result supported the hypothesis that subjective strategy efficacy, as reflected in postdictions, mediates the updated knowledge reflected in post-task strategy effectiveness ratings.

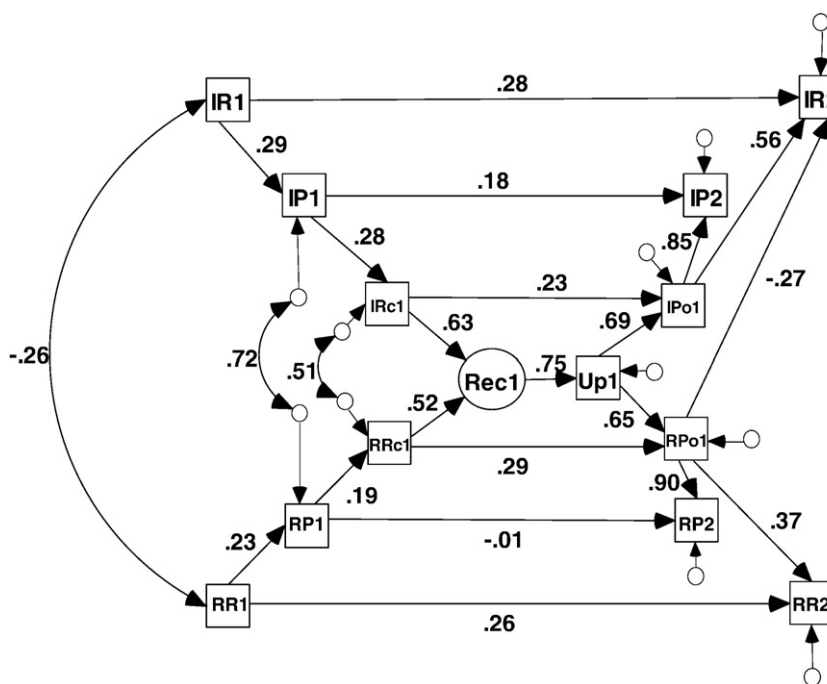


Fig. 3. Structural regression model with standardized regression coefficients (direct effects), Experiment 1.

2.3. Discussion

The results of Experiment 1 conclusively demonstrated knowledge updating. The strategy effectiveness ratings showed large and dramatic shifts in favor of imagery, and these changes were reliably predicted by the differentiated postdictions. Indeed, the evidence in the correlations among measures speaks to profound changes in strategy knowledge as a function of List 1 recall performance experience. There was a slight upgrade in correlations from before to after study on the first list (roughly from 0 to .2), but the bulk of the correlational upgrade in predictions occurred after the List 1 test.

The results of the structural regression models suggest that individual differences in strategy knowledge are determined in large part by accurate performance monitoring. CJs correlated almost perfectly with recall, and very highly with subsequent undifferentiated and differentiated postdictions. The large shifts in the correlations of List 2 predictions with recall were fully mediated by List 1 differentiated postdictions, and these measures of subjective strategy efficacy also mediated the relationship of recall differences between the strategies to knowledge updating, as measured by strategy effectiveness ratings at the end of the experiment. Thus, these data indicate the potency and primacy of performance monitoring as an explanation for knowledge updating.

The fact that recall for each strategy predicted differentiated postdictions is important, for it rules out the hypothesis that a global frequency detection process could produce the knowledge updating. Instead, individuals have access to information about how well they did with each strategy. This knowledge could be due to relatively automatic frequency counts for each strategy or the application of a frequency-informed decision heuristic (e.g., availability; Brown, 1995; Maley et al., 2000).

The high correlation of undifferentiated postdictions with the differentiated postdictions suggests either that the global performance estimate influences the differentiated postdictions directly or that they share a strong, common origination process. Given the problem of equivalent models having equivalent fits (e.g., MacCallum, Wegener, Uchino, & Fabrigar, 1993), we cannot test this hypothesis in the present data set. New variables based on process-oriented models would need to be included in the measurement design.

Another new finding in this study relates to the relationship of pre-experimental strategy knowledge to initial metacognitive predictions. Strategy effectiveness ratings were related to participants' initial global predictions of how well they would perform when using each strategy. This outcome is consistent with metacognitive models in the literature that posit a role for strategy knowledge in initial strategy selection (e.g., Dunlosky & Hertzog, 1998b; Winne, 1996). Moreover, there was some stability of individual differences in effectiveness ratings, but also dramatic shifts reflecting knowledge updating grounded in task experience.

3. Experiment 2

Experiment 2 evaluated in part whether absolute accuracy of the metacognitive judgments could be improved by experimental manipulation (see Hertzog et al., in press). We used blocked testing as a between-subjects experimental manipulation to evaluate improvement in the absolute accuracy of global differentiated postdictions. Indeed, blocked testing did improve the absolute accuracy of these measures. The critical question for the present paper was whether this manipulation would also have an impact on individual differences in knowledge updating effects.

3.1. Method

Details on the experimental design can be found in Hertzog et al. (in press). As before, the PEP I and PEP II were administered before and after the task to assess strategy knowledge updating. All metacognitive judgments collected in Experiment 1 were also collected in Experiment 2, with the exception of CJs. No accuracy feedback was provided.

3.2.1. Participants

The sample consisted of 126 undergraduate students at the Georgia Institute of Technology who received extra credit in psychology courses in exchange for their participation. Individuals were randomly assigned to receive either standard (i.e., random) or blocked testing.

3.2.2. Procedure

Participants completed two study-test trials, each using a different 60-item list of PA. Testing was conducted on a subset of 40 items to maximize the probability of filling the test-set with 20 items reported as studied with each strategy. In the random testing condition, the 40 items were tested in a random order. The blocked testing condition consisted of four blocks of 10 PA items, 2 blocks for each strategy. At the start of each block, participants were informed which strategy they had reported using to encode the items. Strategy feedback was based on reported strategy use and the subset of 40 items was used to address issues of noncompliance with instructed strategies and to increase the likelihood of homogeneous blocks of recall testing for those in the blocked testing condition. Participants complied with imagery instructions 83% (SD=15.2) of the time on List 1 and 84% (SD=17.4) on List 2, with corresponding repetition compliance rates of 67% (SD=24.4) and 63% (SD=26.8). Less than perfect compliance with instructed strategies was also the basis for analyzing the data as a function of reported rather than instructed strategy use.

3.2. Results

3.2.1. Actual and perceived strategy effectiveness

As in Experiment 1, recall levels indicated imagery was a far more effective strategy ($M=.67$, $SD=.22$) than repetition ($M=.28$, $SD=.20$), $d=2.00$. Strategy effectiveness ratings for imagery and repetition were equal before task experience ($M=6.8$, $SD=2.1$), but separated widely after task experience (Imagery $M=8.2$, $SD=1.6$; Repetition $M=3.5$, $SD=2.1$), $d=2.46$.

The blocking manipulation had no effect on recall, but did influence the strategy effectiveness ratings (see Hertzog et al., in press, for details) by both increasing imagery effectiveness ratings and decreasing repetition ratings. These outcomes indicate that participants learned that imagery was the superior strategy after task performance, and this knowledge updating was enhanced by the blocked testing.

3.2.2. Individual differences in metacognitive judgments

We computed correlations between the major metacognitive judgments, recall performance, and strategy effectiveness

Table 3

Pearson correlation coefficients of individuals' predictions, postdictions, and mean JOLs with recall performance in Experiment 2

	List 1		List 2	
	Imagery	Repetition	Imagery	Repetition
Global prestudy predictions	.11	.03	.52**	.54**
Mean JOLs	.22*	.19*	.44**	.46**
Global poststudy predictions	.26**	.24**	.58**	.49**
Global postdictions	.75**	.74**	.69**	.69**

Note. All global judgments refer to global-differentiated judgments. JOLs=Judgments of learning; 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; **= $p<.01$; *= $p<.05$.

ratings. The patterns of the correlations for the measures across lists were a close replication of the data from Experiment 1, so we report only selected correlations to illustrate the general pattern. Table 3 reports the critical correlations for the metacognitive judgments, collapsing over the experimental conditions. Correlations of predictions and mean JOLs with recall were low for List 1, but correlations of List 1 postdictions with performance were much higher, leading to higher List 2 prediction-recall correlations.

Other correlations not reported in Table 3 also replicated key findings from Experiment 1. First, the correlations of List 1 differentiated postdictions with List 2 differentiated predictions were .88 for imagery and .90 for repetition, consistent with the argument that inferences about List 1 strategy effectiveness carried over directly to the differentiated predictions for List 2. Second, the correlations of differentiated predictions across lists were far less stable, .33 and .35 for imagery and repetition, respectively, suggesting that knowledge updating, driven by differentiated postdictions, altered individual differences in predictions. Third, List 1 and List 2 mean JOLs were far more stable than global differentiated predictions, correlating .78 and .77 for imagery and repetition, respectively, showing that JOLs were less likely to change as a function of task experience.

3.2.3. Recall, recall performance estimates, and subjective strategy efficacy

Given the smaller sample size in Experiment 2, the complexity of the model estimated in Experiment 1, and the potent effect of retrieval blocking on judgment accuracy, we opted not to report a structural equation model for these data. Too few students participated to justify assumptions needed for separate structural regression models, especially after dividing the sample into random and blocked testing subgroups (Bollen, 1989).

Instead, we computed selected regression and partial correlation analyses of the data to test our major hypotheses. First, we tested the specificity of the relationship of recall with each strategy to the differentiated postdictions. We computed partial correlations of differentiated postdictions for List 1 with reported recall for each strategy, controlling on the undifferentiated List 1 postdiction. Consistent with the structural regression model for Experiment 1, these partial correlations were significantly different from zero for both strategies—.48

for imagery, and .61 for repetition. This result replicated the finding that differentiated postdictions were influenced by strategy-specific performance monitoring that could not be accounted for by monitoring overall level of recall. Moreover, these partial correlations were higher in the blocked condition than in the random condition. For imagery, the partial correlations were .41 ($df = 58, p < .001$) and .57 ($df = 62, p < .001$) in the random and blocked conditions, respectively. For repetition, the partial correlations were .54 ($df = 58, p < .001$) and .70 ($df = 62, p < .001$). Thus, blocked testing increased the influence of strategy-specific information on inferences about strategy effectiveness, as manifested in the differentiated List 1 postdictions.

3.2.4. Predictors of knowledge updating

We then evaluated the relationship of recall, postdictions, and strategy ratings in multiple regression models. Because the pattern of correlations was similar among these variables for the two lists, we aggregated the data over lists to produce higher reliability of the postdiction variables. We then computed the difference between PEP II imagery and repetition effectiveness ratings as our measure of strategy knowledge at the end of the task, and the difference in differentiated postdictions for imagery and repetition as the operational definition of subjective strategy efficacy. We also computed the aggregate difference in imagery and repetition recall over the two lists. Table 4 reports the correlations of these variables. Note that these data falsify the commonly accepted myth that difference scores cannot correlate with other measures because of low reliability (e.g., Cronbach & Furby, 1970). The differences in strategy effectiveness ratings correlate more highly with differences between the differentiated postdictions for imagery and repetition than the component ratings correlate with differentiated postdictions themselves. Such an outcome lends credence to the difference in effectiveness ratings being viewed as a key measure of strategy knowledge. Note also that the difference in postdictions was more highly correlated with the difference in strategy effectiveness ratings than the actual difference in recall between the strategies (.51 versus .38), supporting the argument that metacognition about performance is the proximal influence on declarative strategy knowledge.

The least-squares regression model showed that this difference in strategy effectiveness ratings was reliably predicted by the difference in imagery and repetition postdictions,

Table 5
Final regression model predicting PEP II rating differences

Source	b (SE)	β	t
Intercept	4.02 (0.48)	–	8.45***
Imagery–Rote Postdiction	0.032 (0.005)	.58	6.79***
Agg.Undifferentiated Postdiction	–0.017 (0.006)	–.23	–2.68**
PEP I Imagery–Rote Ratings	0.182 (0.064)	.21	2.85**

Abbreviation: Agg.—Aggregate.

** $p < .01$.

*** $p < .001$.

controlling on the difference in PEP-I ratings and on the aggregate level of undifferentiated postdictions. All effects were still reliable controlling on the difference in actual recall between imagery and repetition, which was not significant, standardized $\beta = .01, t < 1$. A dummy variable representing blocked versus random testing was not significant, $F < 1$, suggesting that postdictions carried the effect of blocked testing to strategy effectiveness ratings. Moreover, the interaction of test format with the postdiction difference was unreliable, $F < 1$. Table 5 reports the final regression model, which accounted for 34% of the variance in ratings differences, $F(1,122) = 20.44, p < .001$. Clearly, the difference in postdicted performance for imagery and repetition was the dominant influence on change in strategy effectiveness ratings.

3.3. Discussion

Organizing recall tests into homogenous blocks by reported strategy use (either imagery or repetition) repaired much of the poor absolute accuracy for List 1 postdictions and List 2 predictions (for details, see full data analysis in Hertzog et al., in press). The analyses reported here show that blocked testing also impacted the correlations of differentiated postdictions with past performance, strengthening the specific effect of imagery recall or repetition recall on imagery postdictions or repetition postdictions, controlling on the undifferentiated postdiction. This outcome indicates that the blocked testing effect is powerful enough to affect individual differences in subjective strategy efficacy.

Furthermore, the data strongly support the hypothesis that subjective strategy efficacy is directly, if imperfectly, used as the basis for strategy knowledge updating. First, the regression analysis showed that differentiated postdictions related to

Table 4
Correlations of PEP II effectiveness ratings and rating differences with postdictions and recall

Variable	1	2	3	4	5	6	7	8
1. Undifferentiated Postdiction	1							
2. Imagery Postdiction	.88**	1						
3. Repetition Postdiction	.67**	.39**	1					
4. Imagery—Repetition Postdiction	.49**	.80**	–.24**	1				
5. Imagery—Repetition Recall	.14	.39**	–.36**	.65**	1			
6. PEP II Imagery Rating	.16	.20*	–.02	.22*	.06	1		
7. PEP II Repetition Rating	.00	–.25**	.44**	–.54**	–.48**	–.25**	1	
8. PEP II Imagery—Repetition	.08	.28**	–.32**	.51**	.38**	.72**	–.85**	1

Note: ** = $p < .01$; * = $p < .05$.

strategy-specific recall measures above and beyond what could be expected from the undifferentiated postdiction, implicating strategy-specific performance monitoring. Second, there was no effect of recall differences in strategies on the difference in strategy efficacy ratings, supporting the argument that subjective strategy efficacy mediates effects of recall differences on knowledge updating. Third, even though blocking had an impact on postdiction accuracy, there were no additional effects of blocked testing on PEP-II ratings, controlling on the differentiated postdictions. In essence, manipulating the test format had experimental effects on postdictions that carried the experimental effect on the magnitude of strategy knowledge updating. These outcomes therefore provide strong support for the hypothesis that subjective strategy efficacy is the principal influence on strategy knowledge updating.

4. General discussion

Individuals do indeed learn about the differential effectiveness of repetition and imagery as mediational strategies with limited PA task experience. This learning is manifested in (a) substantial changes in imagery and repetition strategy effectiveness ratings, (b) changes in differentiated postdictions after List 1 recall, and (c) subsequent shifts in List 2 differentiated predictions.

The major influence on this updating is accurate performance monitoring, which yields declarative knowledge about strategy effectiveness through inferential mechanisms. As with previous research on postdictions, the undifferentiated postdiction correlated much higher with performance than previous predictions, and the structural equation model for Experiment 1 showed the accuracy of the differentiated postdictions accounted for the substantial increase in correlations of List 2 predictions with recall.

At present we do not know the specific mechanisms that produce the increase in correlations just identified. It could be due to a relatively automatic performance monitoring process, in which performance outcomes are stored and easily accessed at a later time (cf. White & Wixted, 1999). However, on average, individuals underestimated imagery performance by more than 20%. This outcome places important constraints on any automaticity-based view of performance monitoring. In our view this type of underconfidence, combined with strong overall correlations of postdictions with performance, is evidence in favor of heuristic inferential mechanisms operating that, while generally accurate, produce a reliable distortion in estimated strategy effectiveness (Brown, 1995; Manis et al., 1993; Winne, 1996).

It therefore appears the inferences that follow performance monitoring are not routinely accurate when multiple, different outcomes must be associated with the conditions that created them, as is the case with learning about differential strategy effectiveness. In Experiment 1, CJs were almost perfectly correlated with item recall (see also Dunlosky & Hertzog, 2000). But the high accuracy of CJs did not ensure that aggregate postdictions were equally accurate. Indeed, the structural model in Experiment 1 detected a direct effect of subjective strategy efficacy for repetition on the post-experimental imagery

efficacy rating, which would not be expected if the postdictions derived directly from monitoring performance outcomes for each strategy. Thus, individuals appear to have fallible subjective strategy efficacy after performance. They have learned that imagery is superior to repetition, but are seriously misguided about exactly how much better performance with imagery is. In Koriat and Bjork's (2006) terms, there can still be clear distortions in experience-based metacognitive judgments, even when experience has a profound impact on the declarative knowledge that imagery is a superior strategy to repetition.

The present data indicate that learning about strategy effectiveness does not occur during encoding experience, prior to the recall test. JOLs and poststudy predictions were slightly more correlated with performance than initial predictions, but the largest, statistically reliable boost in accuracy only came after test performance. Therefore, monitoring strategy use during study apparently does not contribute greatly to knowledge about differential strategy effectiveness. Instead, performance monitoring and subsequent inferential mechanisms are critical for strategy knowledge updating (Dunlosky & Hertzog, 2000; Finn & Metcalfe, 2007; Koriat & Bjork, 2006).

As noted earlier, declarative knowledge about strategy effectiveness has been shown to predict spontaneous use of effective mediational strategies for PA learning (Hertzog & Dunlosky, 2006). An interesting question for future research is whether declarative knowledge predicts subsequent shifts in strategic choice (e.g., Rabinowitz, Freeman, & Cohen, 1992; Schunn & Reder, 2001). In these experiments we did not transfer individuals to a condition in which they were free to choose any strategy they wished. One would predict, based on the compliance results we reported, that after task experience individuals would be far more likely to choose imagery over repetition if they were free to use any strategy. Rabinowitz et al. (1992) showed that adoption of a strategy is not merely a function of its perceived effectiveness, but also its difficulty of use. Imagery is certainly more difficult to use than repetition, but obviously far more effective. Whether declarative knowledge of its effectiveness outweighs the perceived difficulty of its implementation, leading to a shift to imagery strategy use in an uninstructed context, is an important question.

We do not claim that the present task environment generalizes to others. For instance, self-regulated learning in educational environments is far more complicated and open-ended than the present structured learning environment, and is characterized by context-driven variability in study behavior (Hadwin et al., 2001). Moreover, strategy implementation in PA learning is very discrete and item-specific, whereas in other task environments learning strategies may be complex and involve relational strategies for integrating information that make it more difficult for individuals to easily attribute success or failure to the strategies used during encoding. If anything, more complex learning environments may require more metacognitive self-regulation than the PA task environment (Schraw & Nietfeld, 1998; Winne, 1996). In such cases, motivational variables may well interact with the mechanisms observed here to produce strategy knowledge, and through that knowledge, strategy selection and adaptive self-regulation through

monitoring learning successes and failures (e.g. Pintrich et al., 2000; Winne, 1996).

Nevertheless, at least some of the mechanisms we have studied here appear to be relevant to students' learning about academic strategies that are effective (e.g., Winne & Jamieson-Noel, 2002). In particular, students must learn and update their knowledge about which strategies are likely to maximize learning across a wide variety of courses and domains. Students that do monitor the effectiveness of different learning strategies at encoding and test and can flexibly shift their strategies to meet different course demands are likely to be more successful than students that do not attempt to use or monitor the usefulness of strategies or even realize they should. Thus, many of the processes that contribute to knowledge updating in a PA task are likely also important in educational settings as students attempt to identify the best strategies for learning. Granted, in other task environments, individuals may not even be able to declare what their processing algorithm is, let alone to be strategic about selecting one processing algorithm over another (e.g., Crowley et al., 1997). At this point, however, we believe the present data suggest that awareness and choice are involved in strategic self-regulatory behaviors in the PA learning task (see also Dunlosky & Hertzog, 1998b; Hertzog & Dunlosky, 2004, 2006).

In summary, individuals learn that imagery is superior to repetition as a result of accurate, strategy-specific item-level performance monitoring that occurs at test (not study). The learning that occurs is inferential in nature, and is not perfectly calibrated with actual performance differences between the strategies. This process leads to some distortion in metacognitive judgments, such as postdictions. Item-level judgments, such as JOLs for future recall, are less influenced by strategy knowledge than aggregate or global predictions of performance with each strategy, suggesting that people discount the strategy used when making JOLs. Nevertheless, the subjective strategy efficacy that is reflected in postdictions influences declarative knowledge about strategy effectiveness that apparently lasts beyond the experimental session itself and can influence self-regulatory behavior in other task contexts.

Acknowledgements

This research was supported by NIA Grant R37 AG13148 awarded to Christopher Hertzog.

We would also like to acknowledge the assistance of William J. Frentzel, Simeon Feldstein, and Ailis Burpee who facilitated data collection for the two experiments.

References

- Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, 33, 587–605.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging*, 9, 287–302.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning & Verbal Behavior*, 9, 529–533.
- Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1539–1553.
- Cavanaugh, J. C., Feldman, J. M., & Hertzog, C. (1998). Memory beliefs as social cognition: A reconceptualization of what memory questionnaires assess. *Review of General Psychology*, 2, 48–65.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, 12, 50–71.
- Cronbach, L. J., & Furby, L. (1970). How should we measure “change”—or should we? *Psychological Bulletin*, 74, 68–80.
- Crowley, K. U., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review*, 17, 462–489.
- Devolder, P., & Pressley, M. (1992). Causal attributions and strategy use in relation to memory performance differences in younger and older adults. *Applied Cognitive Psychology*, 6, 629–642.
- 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., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging*, 13, 597–607.
- Dunlosky, J. T., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- 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.
- Fabricsius, W. V., & Cavalier, L. (1989). The role of causal theories about memory in young children's memory strategy choice. *Child Development*, 60, 298–308.
- Fabricsius, W. V., & Hagen, J. W. (1984). Use of causal attributions about recall performance to assess metamemory and predict strategic memory behavior in young children. *Developmental Psychology*, 20, 975–987.
- Finn, B., & 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.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Gigerenzer, G., & Todd, P. M. ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hadwin, A. F., Winne, P. H., Stockley, D. B., Nesbit, J. C., & Woszczynna, C. (2001). Context moderates students' self reports about how they study. *Journal of Educational Psychology*, 93, 477–487.
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. The case of frequency of occurrence. *American Psychologist*, 39, 1372–1388.
- Herrmann, D. J., Grubs, L., Sigmundi, R. A., & Grueneich, R. (1986). Awareness of memory ability before and after relevant memory experience. *Human Learning*, 5, 91–107.
- 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: John Benjamins Publishers.
- Hertzog, C., Dixon, R. A., & Hulstsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, 5, 215–227.

- 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, Neuropsychology, and Cognition, 1*, 203–237.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29*, 22–34.
- Hertzog, C., Price, J., Burpee, A., Frenzel, W.J., Feldstein, S., & Dunlosky, J. (in press). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology*.
- 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.
- Hultsch, D. F., Hertzog, C., Dixon, R. A., & Davidson, H. (1988). Memory self-knowledge and self-efficacy in the aged. In M. L. Howe, & C. J. Brainerd (Eds.), *Cognitive development in adulthood: Progress in cognitive development research* (pp. 65–92). New York: Springer.
- Hunter-Blanks, P., Ghatala, E. S., Pressley, M., & Levin, J. R. (1990). Comparison of monitoring during study and during testing on a sentence-learning task. *Journal of Educational Psychology, 80*, 279–283.
- Jöreskog, K. G., & Sörbom, D. (2003). *LISREL 8.54 user's guide*. Hillsdale, NJ: SSI Scientific Software International.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, U.K.: Cambridge University Press.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology, 126*, 349–370.
- Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learner's sensitivity to retrieval conditions at test. *Memory & Cognition, 34*, 959–972.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgment of learning exhibit increased under-confidence-with-practice. *Journal of Experimental Psychology: General, 131*, 147–162.
- Kuhn, D. (2001). How do people know? *Psychological Science, 12*, 1–8.
- 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.
- Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition, 5*, 264–296.
- Lovelace, E. A. (1990). Aging and metacognitions concerning memory function. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness, and interventions* Oxford, England: North-Holland.
- Lovett, M. C., & Schunn, C. D. (1999). Task representations: Strategy variability, and base-rate neglect. *Journal of Experimental Psychology: General, 128*, 107–130.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185–199.
- Maley, J. E., Hunt, M., & Parr, W. (2000). Set-size and frequency-of-occurrence judgements in young and older adults: The role of the availability heuristic. *Quarterly Journal of Experimental Psychology, 53A*, 247–270.
- Manis, M., Shedler, J., Jonides, J., & Nelson, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality and Social Psychology, 65*, 448–457.
- 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.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
- Nelson, T. O., & Narens, S. L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation, vol. 26*. (pp. 125–173) New York: Academic Press.
- Paivio, A. (1978). Mental comparisons involving abstract attributes. *Memory & Cognition, 6*, 199–208.
- Pearman, A., & Storandt, M. (2005). Self-discipline and self-consciousness predict subjective memory in older adults. *Journal of Gerontology: Psychological Sciences, 60*, P153–P157.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, NE: Buros Institute of Mental Measurements.
- Pressley, M., & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology, 47*, 430–450.
- Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior, 23*, 270–288.
- Rabinowitz, M., Freeman, K., & Cohen, S. (1992). Use and maintenance of strategies: The influence of accessibility to knowledge. *Journal of Educational Psychology, 84*, 211–218.
- Richardson, J. T. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5*, 597–614.
- Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language, 51*, 71–79.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty (2nd ed)*. Mahwah, NJ: Erlbaum.
- Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology, 90*, 236–248.
- 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.
- Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20*, 216–230.
- Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4*, 131–135.
- Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*, 377–397.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.
- West, R. L., & Yassuda, M. S. (2004). Aging and memory control beliefs: Performance in relation to goal setting and memory self-evaluation. *Journal of Gerontology: Psychological Sciences, 59*, P56–P65.
- White, K. G., & Wixted, J. T. (1999). Psychophysics of remembering. *Journal of the Experimental Analysis of Behavior, 71*, 91–113.
- Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8*, 327–353.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*, 551–572.