Metacognitive awareness of learning strategies in undergraduates

Jennifer McCabe

Abstract Two studies examined undergraduates' metacognitive awareness of six empirically-supported learning strategies. Study 1 results overall suggested an inability to predict the learning outcomes of educational scenarios describing the strategies of dual-coding, static-media presentations, low-interest extraneous details, testing, and spacing; there was, however, weak endorsement of the strategy of generating one's own study materials. In addition, an independent measure of metacognitive self-regulation was correlated with scenario performance. Study 2 demonstrated higher prediction accuracy for students who had received targeted instruction on applied memory topics in their psychology courses, and the best performance for those students directly exposed to the original empirical studies from which the scenarios were derived. In sum, this research suggests that undergraduates are largely unaware of several specific strategies that could benefit memory for course information; further, training in applied learning and memory topics has the potential to improve metacognitive judgments in these domains.

Keywords Metacognition · Metamemory · Learning strategies · Undergraduate education

Introduction

Prior research on metacognition and metamemory in college students does not present an optimistic picture,

J. McCabe (✉)
Psychology Department, Goucher College,
1021 Dulaney Valley Road,
Baltimore, MD 21204, USA
e-mail: jennifer.mccabe@goucher.edu

suggesting overconfidence in self-chosen study strategies relative to academic performance (e.g., Winne & Jamieson-Noel, 2002), as well as very low correlations between self-predicted and actual performance on exams and other learning assessments (e.g., Kornell & Bjork, 2007; Kruger & Dunning, 1999), a pattern that may be especially apparent in low-performing students (e.g., Hacker, Bol, Horgan, & Rakow, 2000). This highlights a dilemma for the strategic allocation of attention and time for studying: if students are not accurate at estimating their own learning and knowledge, then they will not be able to make choices about strategies to improve areas that are weakly represented. To compound matters further, a necessary prerequisite for *choosing* a strategy is basic metacognitive *knowledge* about which learning strategies are beneficial for long-term memory. The present research aimed to address this latter issue by examining undergraduates' predictions about learning outcomes, based on educational scenarios derived from published research studies (Study 1), and further, to explore how targeted instruction on applied memory topics relates to scenario prediction accuracy (Study 2).

Undergraduates may not utilize the most effective learning strategies. On an open-ended survey question regarding such strategies, college students most frequently reported "rereading notes or textbook" (Karpicke, Butler, & Roediger, 2009, p. 475). Similarly, Hartlep, and Forsyth (2000) found that most students reported reading and highlighting important concepts, then reviewing whatever they had highlighted. Importantly, in both studies, most students failed to mention a variety of techniques that have been shown to be effective in prior research; when empirically-supported techniques were listed (e.g., self-testing), they were ranked relatively low (Karpicke et al., 2009). As a piece of evidence related to the instruction

issue, 80% of undergraduates in another recent survey reported that their study strategies were improvised, and *not* taught to them in a formal manner (Kornell & Bjork, 2007). This raises the question of whether those improvised strategies, presumably based on intuition and/or metacognitive feedback, are consistent with what is known to be effective from research, and also whether targeted instruction on learning and memory topics could improve metacognitive awareness of successful learning strategies.

Based on recommendations and references from "*25 Learning Principles to Guide Pedagogy and the Design of Learning Environments*" (University of Memphis, 2008) and from "*Organizing Instruction and Study to Improve Student Learning*" (Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel, & Metcalfe, 2007), I chose six learning strategy topics to survey via the Internet. The topics were purposefully chosen to represent effective strategies that were not on the surface intuitive (and indeed, some could be considered counterintuitive). I identified one representative published study for each topic whose methods I could transcribe into a brief scenario-type description. For each scenario, participants rated the predicted level of effectiveness for two contrasting learning situations (one empirically-supported and one not), for typical college students and for themselves. The two ratings were included based on the possibility that they would elicit differing degrees of accuracy. On one hand, it was possible that having to rate other students would enhance objectivity in deciding which scenario might result in the best learning; however, it was also possible that applying the scenario directly to oneself would enhance the depth of appraisal, and therefore accuracy, of predictions.

This research differs from prior work on metacognitive judgments, in that instead of participants directly experiencing the learning conditions, and then making judgments of learning (JOLs) (e.g., Dunlosky & Nelson, 1994), they were instead asked to rate learning outcomes from hypothetical scenarios. Although JOLs are in part determined by mnemonic cues derived from actual exposure to the to-be-learned materials, Koriat's (1997) cue-utilization framework posits that they are also influenced by participants' *a priori* beliefs and theories about which conditions lead to optimal learning outcomes. It is this component of the cue-utilization view that is investigated here. The learning scenarios chosen for this study were presumably evaluated using mainly *extrinsic* cues, based on knowledge of the learning conditions and encoding operations presented in each scenario (e.g., self-generation of materials), as opposed to *intrinsic* cues, based on the nature of the study items themselves (e.g., degree of relatedness), a variable not addressed in the scenario descriptions.

By asking participants to provide JOLs based solely on what they believe to be true about various conditions for learning, and not confounded by intrinsic or real-time mnemonic cues, the extrinsic cue contribution of JOLs can be more effectively isolated. This is important because metacognitive illusions reported in the literature (e.g., Karpicke, 2009; Koriat & Bjork, 2006) presumably arise at least in part from this type of immediate experience, in which people are 'fooled' into thinking they have learned more than they have, perhaps through increased short-term familiarity or fluency ratings (e.g., Benjamin, Bjork, & Schwartz, 1998). The current study therefore makes a unique contribution to the literature by providing a cleaner account of students' *a priori* metacognitive knowledge and beliefs regarding the best ways to learn and remember information. Also unique is the use of concrete descriptions of specific learning scenarios, as opposed to more abstract questions (i.e., "What kind of strategies do you use when you are studying?") used in prior survey research (e.g., Karpicke et al., 2009; Kornell & Bjork, 2007).

I hypothesized that I would overall find low correspondence between students' predictions and research findings on these six topics, which would be consistent with the literature described above. This may be particularly true for the items that are especially counterintuitive (e.g., the advantage of still photos over animated video; Mayer, Hegarty, Mayer, & Campbell, 2005). There are at least two conceptual underpinnings for the prediction about low metacognitive awareness of these strategies. First, although participants do not have the opportunity to experience a real-time metacognitive illusion, they may draw on episodic memories for such an experience, and/or base their responses on potentially biased semantic knowledge (i.e., personal theories of learning and memory). Second, a lack of explicit training in optimal learning strategies, or memory principles in general, could contribute to the failure to endorse optimal learning situations, a point addressed more fully in Study 2.

Below I describe each of the six learning scenario topics (with the empirically-supported outcome listed first), along with a discussion of prior research on college students' metacognition with regard to each topic, when available. By way of preview, the topics related to *cognitive load theory* have not been investigated through the lens of metacognition; however, the latter three topics (i.e., *testing, spacing, generation*) have been a focus of metacognitive research.

The first three learning scenarios were constructed using specific applications of *cognitive load theory* (e.g., Paas, Renkle, & Sweller, 2003), which states that because cognitive resources (i.e., attention and working memory) are limited, effective instruction should reduce the amount of cognitive resources needed to process the to-be-learned material, with the particular goal of reducing extraneous or redundant information. In other words, the extent to which limited cognitive resources can be focused on relevant to-

be-learned information should result in better learning outcomes.

Scenario 1: Dual-code vs. single-code presentations The first survey item was based on *dual code and multimedia effects* (e.g., Mayer & Moreno, 1998, 1999, 2003), which states that memory is better for material presented in multiple modalities (e.g., auditory and verbal), due to an increase in the number and richness of retrieval routes (Paivio, 1986), and also due to the utilization of separate pools of resources in working memory (e.g., Kalyuga, Chandler, & Sweller, 1999). Thus, researchers recommend that learning materials should be designed to make use of multiple modes, taking care to manage the cognitive load of the presentation as to not overwhelm the learner's attention and working memory capacity. The survey item corresponding to this principle was a summary of Kalyuga et al. (1999), in which a significant learning advantage was found for viewing a diagram accompanied by auditory verbal information presented via headphones, in comparison to viewing the diagram accompanied by the same verbal information presented visually as text on the screen.

Scenario 2: Static vs. animated media The second survey item was based on a specific prediction of cognitive load theory called the *static-media hypothesis*. This states that animated visuals (i.e., video) use more cognitive resources than comparable static (or *still*) visuals (e.g., pictures, illustrations), perhaps due to the presentation of more extraneous details (e.g., Mayer et al., 2005). Thus, learning should be better in a static media situation compared to an animated media situation, a pattern found in four experiments by Mayer et al. (2005). For the survey item, I summarized the methods of Mayer et al., portraying learning about a scientific topic from an animated video versus a series of static illustrations.

Scenario 3: Low-interest vs. high-interest details The third survey item was based on using cognitive load theory to understand how the inclusion of *extraneous (irrelevant) details* affects learning. Mayer, Griffith, Jurkowitz, and Rothman (2008) demonstrated that high-interest extraneous details included in a slideshow presentation on a scientific topic led to poorer learning outcomes than low-interest details. High-interest details were hypothesized to use more of the limited processing resources available to the learner, with the result that the learner had fewer resources to focus on the to-be-learned information. The survey item included a description of slideshows that presented high-interest versus low-interest extraneous details.

The next three learning scenario survey items were relevant to other empirically-supported cognitive principles

in education: the *testing effect,* the *spacing effect,* and the *generation effect*. These are all examples of effortful yet memory-enhancing strategies coined *desirable difficulties* by Bjork (1994).

Scenario 4: Testing vs. restudying The fourth scenario was based on the *testing effect*, which states that learning and memory for material is improved when time is spent taking a test on the material, versus spending the same amount of time restudying the material (e.g., Butler, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006a). This phenomenon occurs because taking a test forces the learner to use memory retrieval, which in turn helps to solidify the information in long-term memory. Roediger and Karpicke (2006b, Experiment 1) demonstrated a large advantage of taking a free recall test for a previously studied prose passage, in comparison to spending a comparable amount of time restudying the passage. The survey item in the current study described these two competing strategies.

The basis for hypothesizing that participants in the current study would erroneously predict that restudying is superior to testing came from several studies. First, participants who received more 'study' than 'test' sessions with a prose passage were more confident in their 1-week-delayed recall, even though those with more 'test' sessions performed better (Roediger & Karpicke, 2006b, Experiment 2). Similarly, Kornell and Son (2009) showed that even when students perform better after self-testing using flashcards, they still display the metacognitive error of choosing restudying over testing (see also Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Karpicke, 2009).

Karpicke et al. (2009) examined the self-testing strategy using survey methodology. Self-testing was not frequently reported as a favorite learning strategy; in addition, when asked about self-testing versus restudying in a forced-choice item, only a minority chose testing as the preferred option. In contrast, a survey by Kornell and Bjork (2007) showed a majority of students endorsed self-testing as a study strategy, but mainly to find out how much they learned; only 18% viewed self-testing as a means to improve memory. In sum, whether or not students actually use self-testing when studying, it is clear that most do not see retrieval practice itself as leading to better learning and subsequent exam performance.

Scenario 5: Spacing vs. massing The fifth survey item was based on the *spacing effect*, which states that, holding constant total study time, spacing out (or distributing over time) the study of to-be-learned material is more effective than massing (or cramming) the material (e.g., Rohrer & Pashler, 2007). Spacing requires more

distinctive retrieval episodes, and may help in discriminating among various problem types (e.g., Kornell & Bjork, 2008). The survey item in the current study was a summary of the methods of Kornell and Bjork, who found that interleaving various artists' paintings during a learning session was more effective for an induction task (i.e., identifying new paintings by the studied artists) compared to grouping the paintings by a single artist together. As such, the study chosen for this scenario, though certainly illustrative of the spacing effect and interpreted as such by the authors, more specifically tested conditions of interleaving versus blocking of materials during study, a topic of recent research interest (e.g., Kornell, Castel, Eich, & Bjork, 2010; Richland, Bjork, Finley, & Linn, 2005).

A metacognitive component of the Kornell and Bjork (2008) study suggested that spacing is not a strategy generally perceived to be effective for memory. When participants were asked which condition led to better memory for the paintings, a minority endorsed spacing, even though a majority performed better on the test in the spacing condition (see also Zechmeister & Shaughnessy, 1980). Other research has been mixed with regard to whether people explicitly choose a massing over a spacing strategy during study (e.g., Benjamin & Bird, 2006; Pyc & Dunlosky, 2010; Son, 2004; Son, 2010; Son & Kornell, 2009; see Kornell & Bjork, 2007 for a review).

Scenario 6: Generating vs. non-generating The sixth survey item was based on the *generation effect*, which states that learner-created materials will be more easily remembered than instructor-provided materials (e.g., DeWinstanley & Bjork, 2004; Schwartz & Metcalfe, 1992). Generating one's own materials is a more difficult and attention-demanding task, and is therefore thought to result in a stronger memory trace for the information. The survey item was a summary of the methods of Bloom and Lamkin (2006), who compared memory for the cranial nerves in instructor-provided and student-generated mnemonic conditions, and found improved memory in the generation condition.

Metacognitive aspects of the generation effect have not been a focus of recent research; however, Begg, Vinski, Frankovich, and Holgate (1991) showed that undergraduates believed generated items would be better recalled than items simply read. Further, judgments of learning are more accurate for generated versus read items (Maki, Foley, Kajer, Thompson, & Willert, 1990; Mazzoni & Nelson, 1995). Therefore, I predicted relatively accurate predictions for the *generation effect* scenario.

Metacognitive self-regulation In addition to the learning scenario ratings, participants self-reported information for a measure of *metacognitive self-regulation*. Metacognition, the ability to evaluate one's own learning, is considered to be an important component of self-regulated learning (e.g., Pintrich, 2000), and is associated with academic monitoring, strategy use, and motivation (e.g., Sperling, Howard, Staley, & DuBois, 2004). In the current study, I examined the correlation between metacognitive self-regulation and scenario prediction accuracy, predicting a positive relationship between the two measures. That is, students who are better able to judge their own learning, and make adjustments accordingly, should show better accuracy when making predictions about the scenarios.

Study 1

The goals of this study were to examine, in a broad sample of undergraduates from a variety of institutions, (1) the accuracy of predictions with regard to the six learning scenarios, (2) the extent to which responses were similar when predicting outcomes for 'yourself' versus for 'typical college students', and (3) correlations between scenario predictions and metacognitive self-regulation.

Method

Participants Participants were 255 current undergraduate students over the age of 18 and with access to the Internet. They were offered the chance at winning a $25 gift card for completion of the survey. Participants were recruited via e-mail and web postings from a variety of sources and institutions of higher education.

Descriptive statistics of demographic variables show that the sample was 78% female and 22% male. The mean age was 24.56 (SD 7.94; median 22; range 18–61). The education level of the sample was overall high, with a mean of 3.72 years of college completed ($SD = 1.33$). Of those who self-reported their current GPA (87%; $n = 222$), the mean was 3.25 ($SD = 0.54$). Participants were asked to report their major; for the purposes of this study, these responses were divided into a category representing psychology (and/or behavioral science) and a category representing all other responses (i.e., non-psychology). Fifty percent of participants reported a psychology major, and 50% reported non-psychology. Exploratory analyses failed to find any difference between psychology majors and non-majors in any of the survey variables (all ps > 0.05). Regarding the number of psychology courses taken by participants the mean was 4.36 ($SD = 3.36$). This was correlated with four survey variables (ps < 0.05); however, because the correlation pattern was inconsistent (i.e., positive in two cases and negative in two), these relationships are not discussed further.

Materials The web-based survey was created by the author and made available to participants via password-protected *SurveyMonkey* software. The final survey consisted of 25 items: six were descriptions of learning scenarios (each followed by 2 rating scales, for 'typical college students' and 'yourself'; see above for descriptions of scenarios), 12 were items from the *Motivated Strategies for Learning Questionnaire* (MSLQ) subscale for *metacognitive self-regulation* (Duncan & McKeachie, 2005; Pintrich, Smith, Garcia, & McKeachie, 1991)*,* and seven were demographic questions.

Learning scenario descriptions were carefully constructed to avoid using the terminology representing the concepts under investigation (e.g., *spacing, generation*), as the goal was to determine if students could predict the outcome by interpreting the given scenario, not by drawing on memory for terms they may have learned about in psychology courses. For example, the survey item for *Scenario 6: Generating vs. non-generating* stated:

> *Two assignments ask students to learn the list of cranial nerves using a mnemonic device. Assignment A includes a commonly used mnemonic device PROVIDED by the instructor to assist students in their learning. Assignment B asks students to CRE-ATE their own mnemonic device to assist their learning. After 2 weeks, all students are asked to list the cranial nerves in order.*

For each item, participants rated on a 7-point scale the extent to which they predicted the two contrasting situations would or would not benefit learning, as measured by subsequent test scores. A rating of "1" would indicate that Situation A should result in much higher test scores, a neutral rating of "4" would indicate that both situations would result in equivalent test scores, and a rating of "7" would indicate that Situation B should result in much higher test scores. To make consistent the Likert-scale data interpretation, prior to analyses I reverse-scored three of the items such that for each scenario result discussed below, higher numbers (i.e., ≥5) indicate endorsement of the conclusion derived from the relevant research study and lower numbers (i.e., ≤3) indicate endorsement of the opposite outcome.

The *metacognitive self-regulation* (MSR) scale (Pintrich et al., 1991; Duncan & McKeachie, 2005) consisted of 12 Likert-type items on a 7-point scale (with "1" corresponding to "not at all true of me" and "7" corresponding to "very true of me"). Examples of statements from the scale include, "If course readings are difficult to understand, I change the way I read the material," "I ask myself questions to make sure I understand the material I have been staying in this class," "I try to think through a topic and decide what I am

supposed to learn from it rather than just reading it over when studying for the course," and "When I study for this class, I set goals for myself in order to direct my activities in each study period." After reverse-scoring two of the items, the mean ratings were computed to form the composite measure of MSR.

Procedure Data collection took place at a participant-chosen day and time, using a link to the survey sent via e-mail. Completion of the survey took between 15 and 30 minutes. Participants first read and agreed to an informed consent form and then were given instructions for participating. They then proceeded to the survey and completed it at their own pace.

Results

For all analyses below, the alpha level was 0.05. One-sample *t*-tests were conducted to compare the mean rating to the neutral response of "4" (i.e., prediction of similar test scores for both situations); in addition, paired-samples *t*-tests were conducted to compare the 'students' and 'yourself' ratings. Original scale ratings were also converted to a percentage of 'correct' endorsement by re-coding responses of 5 and higher as "correct" and other responses as "incorrect." See Table 1 for all descriptive statistics.

Scenario 1: Dual code vs. single code presentations For 'typical college students,' the mean rating of predicted learning outcome (i.e., test performance) was not significantly different from the neutral "4", $t(254) = 1.11$, $p = 0.27$. For 'yourself,' the mean rating was significantly lower than "4," $t(253) = 2.37$, $p = 0.02$. The means for 'students' and 'yourself' were significantly different, $t(253) = 4.24$, $p < 0.001$, with lower ratings for those relevant to the participants themselves. To preview, this difference in ratings was found only for this scenario. The percentage of participants endorsing the empirically-supported outcome was not different for 'students' and 'yourself', $t(253) = 0.46$, $p = 0.65$. To summarize results for this item, participants overall did not endorse the empirically documented advantage for auditory over visual presentation of supplemental information (Kalyuga et al., 1999), and, interestingly, their ratings veered even further toward the lower end of the scale, representing a prediction of a single-code advantage, when they were asked to predict their own learning outcomes.

Scenario 2: Static vs. animated media The mean rating of predicted test outcomes for 'students' was significantly lower than the neutral response of "4," $t(254) = 12.46$, $p < 0.001$, as was the case for the 'yourself' ratings, $t(254) =$

Table 1 Predicted learning scenario outcomes in Study 1

Scenario	"Typical college students"			"Yourself"		
	M	SD	% Correct endorsement	M	SD	% Correct endorsement
1. Dual-code vs. single-code	4.11	1.69	39.22%	3.67*	2.20	37.80%
2. Static vs. animated media	2.84**	1.49	12.16%	2.90**	1.92	21.96%
3. Low- vs. high-interest details	2.55**	1.77	14.96%	2.61**	1.72	14.90%
4. Testing vs. restudying	3.38**	1.97	29.92%	3.37**	2.14	32.68%
5. Spacing vs. massing	2.00**	1.33	6.67%	1.90**	1.48	9.80%
6. Generating vs. non-generating	4.21	1.86	47.45%	4.39**	2.19	53.33%

M mean, *SD* standard deviation, *% Correct Endorsement* is percentage of participants providing a rating of "5" or higher

*$p < .05$; **$p < .01$, for one-sample *t*-tests against neutral value of "4" on survey items. Numbers above "4" indicate endorsement of the empirically supported outcome

9.16, $p < 0.001$. Means for 'students' and 'yourself' were not significantly different, $t(254) = 0.54$, $p = 0.59$. The percentage of participants endorsing the empirically-supported outcome was significantly higher for 'yourself', $t(254) = -4.02$, $p < 0.001$. This set of results suggests that students overall believed that they, and other students, would learn best from an animated science presentation, a prediction that is opposite of the Mayer et al.'s (2005) results. However, there was a higher rate of 'accuracy' in predictions for 'yourself' compared to 'students' ratings, a pattern found only in this item.

Scenario 3: Low-interest vs. high-interest details Results showed that the mean rating for 'students' was significantly lower than the neutral response, $t(253) = 13.06$, $p < 0.001$; a similar pattern was found for 'yourself', $t(254) = 12.93$, $p < 0.001$. The means for 'students' and 'yourself' were not significantly different, $t(253) = 0.84$, $p = 0.40$. Regarding percentage of correct endorsement, the means were identical, $t(253) = 0.00$, $p = 1.00$. This set of results indicates that, contrary to research findings (Mayer et al., 2008), students believed that better learning would result from the inclusion of high-interest extraneous details in a presentation.

Scenario 4: Testing vs. restudying The mean rating for this scenario for 'students' was lower than the neutral "4" response, $t(253) = 5.00$, $p < 0.001$, as was the rating for 'yourself', $t(253) = 4.66$, $p < 0.001$. The means for 'students' and 'yourself' were not different, $t(252) = 0.05$, $p = 0.96$. The mean percent correct was similar for the two ratings, $t(252) = -1.27$, $p = 0.21$. Thus, most students perceived a learning advantage for restudying material over taking a retrieval test, in stark contrast to empirical research showing a benefit for testing over restudying on subsequent test performance (Roediger & Karpicke, 2006b).

Scenario 5: Spacing vs. massing The mean for 'students' was significantly lower than the neutral response, $t(254) = 23.97$, $p < 0.001$; and the mean for 'yourself' was also significantly lower than neutral, $t(254) = 22.66$, $p < 0.001$. The means for 'students' and 'yourself' were not significantly different, $t(254) = 1.53$, $p = 0.13$. Mean percent correct endorsement was similar for the two ratings, though slightly higher for 'yourself', $t(254) = -1.90$, $p = 0.06$. Overall, participants showed a clear endorsement of massing over spacing for predicted learning outcomes, contrary to consistent findings in the literature of a benefit for spacing (Kornell & Bjork, 2008).

Scenario 6: Generating vs. non-generating The mean rating for 'students' was not significantly different from the neutral response, but did show a trend to be higher, $t(254) = 1.82$, $p = 0.07$; for 'yourself', the mean was significantly higher than neutral, $t(254) = 2.86$, $p = 0.005$. The means for 'students' and 'yourself' were not significantly different, $t(254) = 1.51$, $p = 0.13$. There were similar percentages of endorsement of spacing over massing for 'students' and 'yourself', $t(254) = -1.74$, $p = 0.08$. Overall, there was some endorsement for self-generating as a superior study strategy compared to reading material provided by another person (consistent with the results of Bloom & Lamkin, 2006), particularly when participants rated their prediction for themselves.

Global scenario performance As a global measure of metacognitive accuracy on the learning scenario items, I computed for each participant (using the mean of the 'typical college students' and 'yourself' ratings) the percentage of scenarios to which he or she responded with endorsement of the empirically-supported outcome. The overall mean for this variable was 23% (i.e., 1.38 scenarios out of six), indicating that participants were overall not very accurate in their predictions of learning outcomes. In fact,

the highest proportion score for any individual participant was 67% (four scenarios), and 18% of the sample responded with the non-empirically-supported outcome for all six learning scenarios (i.e., 0% accuracy). These descriptive statistics support the original hypothesis that participants would be generally poor predictors of empirically-supported learning outcomes.

Metacognitive self-regulation On the 7-point Likert scale, with higher numbers indicating a higher amount of MSR, the mean score was 4.55 (*SD* = 0.91).

A bivariate correlational analysis indicated that MSR was positively correlated with two demographic variables: GPA, $r(220) = 0.20$, $p = 0.003$, and age, $r(251) = 0.17$, $p = 0.01$. Thus, higher-performing students, and older students, tended to exhibit higher levels of metacognition. With regard to correlations with learning scenario variables, there was a significant relationship between MSR and mean percent correct endorsement for *Scenario 2: Static vs. animated media*, $r(253) = 0.16$, $p = 0.01$, a correlation that remained significant after partialing out education level and age ($ps < 0.05$) and marginally significant after controlling for self-reported GPA ($p = 0.06$). MSR was further correlated with two measures associated with the 'yourself' ratings for *Scenario 6: Generating vs. non-generating*: scale ratings, $r(253) = 0.15$, $p = 0.02$, and percentage of correct endorsement, $r(253) = 0.16$, $p = 0.01$. These correlations remained significant after separately partialing out education level, self-reported GPA, and age ($ps < 0.05$).

Consistent with the original hypothesis, there was a relatively small but significant positive correlation between MSR and the global scenario accuracy percentage, $r(253) = 0.13$, $p = 0.03$, suggesting that students who self-reported stronger endorsement of statements relating to their own metacognition in a classroom setting were overall more accurate in predicting learning outcomes. This correlation remained even after partialing out education level ($p = 0.04$), but was reduced to a trend when partialing out GPA ($p = 0.10$).

Discussion

Out of six learning scenarios tested in Study 1, each containing two rating scales, only one (i.e., the 'yourself' rating for *Scenario 6: Generating vs. non-generating*) showed a mean rating that was significantly different from neutral in the direction of endorsing the research-based finding (i.e., Bloom & Lamkin, 2006), as well as a percentage of correct endorsement over 50%. This is consistent with research suggesting that students have metacognitive awareness of a memory advantage for self-generated information (e.g., Begg et al., 1991). For several other scenarios, however, participants in the current study consistently endorsed the non-empirically-supported outcome, suggesting little or no awareness of which of the presented strategies would be most beneficial for learning.

This pattern of results is not particularly surprising, considering some of the scenarios were based on relatively counterintuitive research findings. Indeed, interpretation of these findings is necessarily limited to the six specific learning scenarios included in the study: college students appear to be overall unaware of the memorial benefits of static media, low-interest extraneous details, testing, and spacing. It is entirely possible, even plausible, that scenarios chosen for their more obvious learning benefits would have resulted in more endorsement of empirically-supported outcomes.

Turning to a comparison of 'yourself' and 'students' ratings, these were similar for all scenarios, with two exceptions: stronger endorsement of the non-empirically-supported outcome for 'yourself' ratings in *Scenario 1: Dual-code vs. single-code*, and higher percentage of correct endorsement for 'yourself' ratings in *Scenario 2: Static vs. animated media*. There is no clear explanation for the two significant findings, and they are opposite in direction; however, I speculate for the former finding that familiarity with the catchphrase "visual learner" could lead those who believe they fit into this category to endorse the fully visual (i.e., single-code) scenario. These results taken together led to the decision for Study 2 to focus on the mean of 'students' and 'yourself' ratings when comparing instruction groups.

MSR was correlated with two scenario variables. Most interestingly, in the case of *Scenario 6: Generating vs. non-generating*, this is also the scenario whose mean rating showed the most agreement with prior research findings (Bloom & Lamkin, 2006), suggesting that the more metacognitively sophisticated students, as measured by the MSR scale, were more likely to predict an advantage for self-generated materials. Given the likely possibility that the generation effect scenario was the most intuitive of the six presented, and was also the only one to individually correlate with MSR, the use of other, more obvious, learning strategies could have led to stronger MSR correlations. These relationships should be investigated in future research.

In support of the original hypothesis, MSR was significantly and positively correlated with the global measure of metacognitive accuracy from the learning scenario rating scales discussed above, a relationship which remained significant after controlling for education level, and remained marginally significant after partialing out GPA. This finding helps to unify patterns of responding from the six scenarios, suggesting that individual differences in metacognition are overall predictive of accuracy in scenario ratings.

Study 2

Study 1 established that for at least five of six learning scenarios derived from published teaching and learning research, college students were unable to predict the results. The purpose of Study 2 was to compare scenario prediction accuracy among several groups of undergraduates who had experienced varying levels of explicit instruction on topics related to the survey items. If instruction on applied learning and memory topics is associated with increased metacognition and subsequent academic performance, as suggested by prior research (e.g., Azevedo & Cromley, 2004; Fleming, 2002; Tuckman, 2003), then participants with more in-depth instruction should perform better on the scenario survey items.

Using the same survey as in Study 1, I tested four groups of students. First, a control group of introductory psychology students who had not received targeted instruction on learning and memory topics was included, with the expectation that they would perform similarly to Study 1 participants. Next, students who were specifically instructed about three of the scenario topics (i.e., those included in the concept of *desirable difficulties*, Bjork, 1994: *testing*, *spacing*, *generation*) were recruited from an introductory psychology course, and also from two sophomore-level cognition courses. Although these students learned the basic cognitive principles listed above, they were not specifically instructed about the research studies used in the learning scenarios. Further, the cognition course students received more in-depth instruction about the topics compared to the introductory psychology students (i.e., spent more class time on, and had more discussion about, the relevant topics). I assessed the extent to which these instruction groups would apply the principles learned in class to the real-world educational situations presented in the survey, and would therefore outperform the non-instruction groups (from Study 1 and Study 2) on the three scenarios related to the topics of targeted instruction. The final group was enrolled in an upper-level seminar course on *Cognition and Education*; throughout the semester, they read and discussed all six articles from which the specific learning scenarios were derived. I therefore expected these students to have high performance on all scenarios.

Method

Participants Participants were undergraduates enrolled in psychology courses at Goucher College. All were offered course credit in exchange for participation. Four groups of participants were tested, representing different levels of specific instruction on topics related to a subset of the learning scenarios. Participants who received instruction were taught by the author. The groups are described below

in approximate order of depth of instruction, from least to most.

Two groups of *Introduction to Psychology* (IP) students were included. The first group consisted of 12 students who did not have any targeted instruction on learning and memory topics. The second group consisted of 50 students who heard a lecture by the author entitled, "*Psychology Principles in Action: Improving Academic Performance*," which included a discussion of ten topics related to applications of cognitive psychology research to education. Those relevant to the current study were testing, spacing, and generation effects. These students participated in the survey approximately 2 weeks after the in-class lecture.

For the Non-Lecture IP group, mean age was 19.25 (SD = 1.22), mean years of college completed was 1.42 (SD = 0.090), and mean self-reported GPA was 3.06 (SD = 0.35). In addition, the sample was 100% non-psychology majors, and 67% female. For the Post-Lecture IP group, mean age was 18.84 (SD = 1.04), mean years of college completed was 1.50 (SD = 0.93), and mean self-reported GPA was 3.18 (SD = 0.51). Ninety-eight percent were non-psychology majors, and 74% were female. For the age, college years, and GPA variables, there were no significant differences between the two IP groups (all ps > 0.05).

The third group of participants consisted of 54 students enrolled in 200-level Cognition courses (*Cognitive Psychology, Human Learning and Memory*) who had learned about education-relevant memory topics (e.g., testing, spacing, generation) in the context of their course(s). Mean age was 20.22 (SD = 1.08), mean years of college completed was 2.35 (SD = 0.93), and mean self-reported GPA was 3.23 (SD = 0.49). Eighty percent were psychology majors, and 91% were female. Cognition course students completed the survey approximately 2 weeks after the last of the survey-relevant topics had been discussed in class.

For the fourth group, 12 students enrolled in an advanced seminar on the topic of *Cognition and Education* were included. Throughout the course, these students had read and discussed the specific research articles from which the six learning scenarios were derived. Mean age was 22.67 (SD = 5.23), mean years of college completed was 3.83 (SD = 0.94), and mean self-reported GPA was 3.51 (SD = 0.39). The sample included 92% psychology majors, and was 83% female. The Seminar students completed the survey near the end of the semester.

Regarding equivalence of the groups on demographic measures, there were no differences for self-reported GPA, $F(3, 109)$ = 1.90, p = 0.13. Not surprisingly, both the Cognition course and Seminar students had higher means for number of college years completed compared to the Post-Lecture IP group; and the Seminar students had more years of education than all other groups (ps < 0.05), $F(3, 124)$ = 24.23, p < 0.001. Also, the Post-Lecture IP group

was younger on average than the Cognition and Seminar groups; and the Seminar students were older than all of the other groups (all $ps < 0.05$), $F(3, 124) = 15.14$, $p < 0.001$. These differences were not assessed further, as Study 1 indicated no relationships between scenario performance and these demographic variables.

Materials and procedure The web-based survey described in Study 1 was used. Details of the procedure were identical.

Results

Preliminary analyses of the 'yourself' and 'students' ratings for the entire group of participants showed that in only one case for each dependent variable did the two ratings significantly differ, i.e., for *Scenario 1: Dual-code vs. single-code* using scale ratings, and for *Scenario 2: Static vs. animated media* using percentage of correct endorsement. Mimicking the pattern found in Study 1, participants gave lower ratings for 'yourself' in *Scenario 1*, and had stronger endorsement of the empirically-supported outcome for 'yourself' in *Scenario 2*. As noted above, for simplicity when presenting the comparisons across groups below, I report results based on the mean of 'yourself' and 'students'[1].

For each scenario, I first describe which groups were significantly above or below the neutral "4" response for scale ratings. I then present analyses comparing all five groups of participants (Study 1 plus the four groups from Study 2) on 7-point scale ratings (see Fig. 1 for all group means) and on percentage of correct endorsement of the empirically-supported outcomes (see Fig. 2 for all group means), to determine whether the group(s) with more in-depth instruction on relevant topics would perform better on scenario items. The Tukey HSD adjustment for multiple tests was applied to all follow-up contrasts.

Scenario 1: Dual code vs. single code presentations For this item, the only group with ratings significantly different than the neutral response of "4" was the Post-Lecture IP group ($M = 3.50$, $SD = 1.52$), $p = 0.024$.

Comparing the five groups on scale ratings using analysis of variance (ANOVA) resulted in no significant differences, $F(4, 378) = 1.07$, $p = 0.37$. The parallel ANOVA for percentage of correct endorsement was marginally significant, $F(4, 378) = 2.40$, $p = 0.05$. The only trend resulting from the follow-up tests was for the Seminar students to outperform the Post-Lecture IP students ($p = 0.083$).

Scenario 2: Static vs. animated media The mean ratings of two groups were significantly lower than neutral: Post-Lecture IP ($M = 3.11$, $SD = 1.44$), $p < 0.001$, and Cognition courses ($M = 2.91$, $SD = 1.45$), $p < 0.001$.

The ANOVA comparing groups on scale ratings was significant, $F(4, 378) = 3.69$, $p = 0.006$, driven by higher ratings in the Seminar group compared to each of the other groups (all $ps < 0.05$), except for the Non-Lecture IP students. Comparing groups on percent correct endorsement was also significant, $F(4, 378) = 3.56$, $p = 0.007$, with an identical pattern of contrasts to that described above.

Scenario 3: Low-interest vs. high-interest details Examining the rating scale means in reference to the neutral "4" response, the Non-Lecture IP group ($M = 2.54$, $SD = 1.63$), $p = 0.01$, the Post-Lecture IP group ($M = 2.90$, $SD = 1.82$), $p < 0.001$, and the Cognition courses group ($M = 2.73$, $SD = 1.81$), $p < 0.001$, were significantly lower than "4".
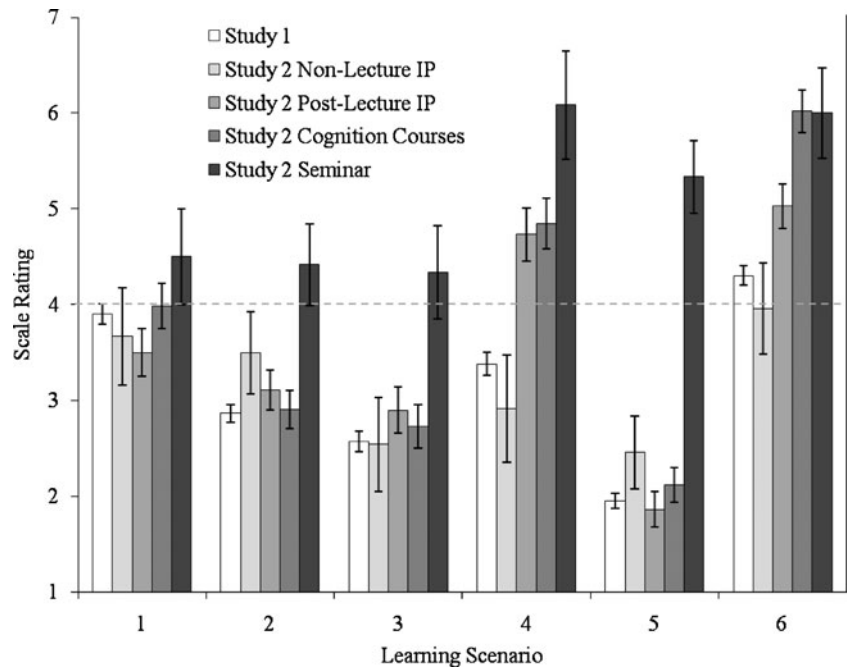
There were significant differences among groups for scale ratings, $F(4, 378) = 3.39$, $p = 0.01$. Here, the Seminar students gave higher ratings than Study 1 participants as well as the Cognition course students ($ps < 0.05$); and there were trends for the Seminar course to also have an advantage over the other two groups ($ps < 0.08$). Groups were also different in the ANOVA using percent correct endorsement, $F(4, 378) = 5.12$, $p < 0.001$, driven by Seminar students outperforming all other groups ($ps < 0.01$).

Scenario 4: Testing vs. restudying For this scenario, only the Non-Lecture IP group was significantly lower than the neutral "4" ($M = 2.92$, $SD = 1.68$), $p = 0.047$; however, the three targeted instruction groups were significantly higher than "4": Post-Lecture IP ($M = 4.73$, $SD = 1.93$), $p = 0.01$; Cognition courses ($M = 4.84$, $SD = 2.14$), $p = 0.004$; and Seminar ($M = 6.08$, $SD = 0.90$), $p < 0.001$.

The ANOVA using scale ratings showed significant group differences, $F(4, 378) = 14.40$, $p < 0.001$. Follow-up contrasts showed that all three groups who learned about the testing effect in class gave higher ratings than Study 1 participants and the Non-Lecture IP students ($ps < 0.05$). No other differences were significant. The parallel ANOVA using percent correct endorsement was also significant, $F(4, 378) = 16.05$, $p < 0.001$, with a pattern of mean contrasts identical to that reported above.

Scenario 5: Spacing vs. massing All non-Seminar groups had ratings significantly lower than "4": Non-Lecture IP ($M = 2.46$, $SD = 1.59$), $p = 0.006$; Post-Lecture IP ($M = 4.73$, $SD = 1.93$), $p < 0.001$; and Cognition courses ($M = 4.84$, $SD = 2.14$), $p < 0.001$. The Seminar students, however, had a mean scale rating significantly higher than "4" ($M = 5.33$, $SD = 1.59$), $p = 0.014$.

---

[1] Major results and conclusions were similar when conducting the between-groups analyses using 'typical college students', 'yourself', or mean ratings.
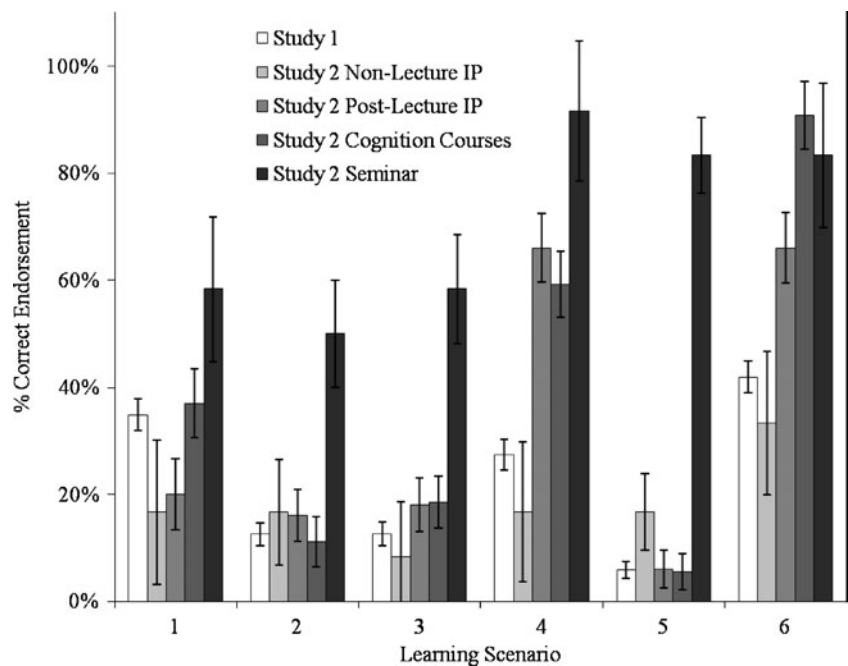
Scale ratings showed significant group differences, $F(4, 378) = 19.74$, $p < 0.001$, driven by higher ratings in the Seminar group compared to all other groups ($ps < 0.001$), who were themselves similar. The ANOVA using percentage of correct endorsement was also significant, $F(4, 378) = 28.85$, $p < 0.001$, with a pattern of contrasts identical to that reported above.

Scenario 6: Generating vs. non-generating Compared to the neutral rating of "4", the three targeted instruction

groups had significantly higher means: Post-Lecture IP, $M = 5.03$, $SD = 1.33$, $p < 0.001$; Cognition coures, $M = 6.02$, $SD = 1.09$, $p < 0.001$; and Seminar, $M = 6.00$, $SD = 0.93$, $p < 0.001$.

The ANOVA comparing groups on scale ratings was significant, $F(4, 378) = 15.47$, $p < 0.001$. Study 1 participants gave lower ratings than all three groups who received targeted instruction in Study 2, and the Non-Lecture IP group was lower than the Cognition and Seminar groups ($ps < 0.05$). Further, students in the

Cognition courses were higher than all groups ($ps < 0.05$) except the Seminar. The parallel analysis using percent correct endorsement also showed group differences, $F(4, 378) = 15.19$, $p < 0.001$, driven by an advantage for all three instruction groups over Study 1 participants, and an advantage of the Cognition group over the Non-Lecture IP students ($ps < 0.05$).

Global scenario performance An ANOVA comparing the five groups on the global measure of scenario performance, computed by averaging the percentage of correct endorsement across the six scenarios, was significant, $F(4, 378) = 33.73$, $p < 0.001$ (see Fig. 3). Follow-up contrasts showed an advantage of all three groups receiving targeted instruction (Post-Lecture IP, 32%; Cognition courses, 37%; Seminar, 71%) over Study 1 (23%) participants. There was an advantage for Cognition and Seminar students over Non-Lecture IP students (18%), $ps < 0.01$, and a trend for the Post-Lecture IP group to outperform the Non-Lecture IP group ($p = 0.06$). The Seminar students outperformed all groups ($ps < 0.001$) on this global measure, with an average of 4.3 scenarios correct out of six.

Metacognitive self-regulation Numerically, the MSR scores paralleled the level of education and targeted instruction (Non-Lecture IP, $M = 4.17$, $M = 0.99$; Post-Lecture IP, $M = 4.25$, $SD = 0.77$; Cognition courses, $M = 4.28$, $SD = 0.72$; Seminar, $M = 4.38$, $SD = 0.68$); however, the means were not significantly different, $F(3, 124) = 0.16$, $p = 0.92$.

Unlike in Study 1, there were no significant correlations of MSR with demographic variables, nor with any scenario performance variables (all $ps > 0.05$). This pattern was evident within each group, and also when combining all participants into one group.

Discussion

As three of the learning scenarios were not targeted for instruction in the non-Seminar courses, it was not surprising that these items (i.e., Scenarios 1, 2, and 3) failed to show any consistent group differences, except for the Seminar group numerically (and at times significantly) outperforming the others. Also, in line with predictions, the control group (i.e., Non-Lecture IP) was statistically similar in all scenario ratings to participants from Study 1.

The main focus of this study was on the latter three scenarios, based on the testing effect, spacing effect, and generation effect. Two survey items, *Scenario 4: Testing vs. restudying* and *Scenario 6: Generating vs. non-generating*, showed similar patterns, with the three groups receiving targeted instruction on the topics outperforming those who

did not[2]. This pattern suggests that instructed students were able to apply basic knowledge about a memory concept (e.g., that retrieval practice provides a mnemonic advantage) to a real-world learning scenario (e.g., that students who take a recall test will outperform those who use the same amount of time to restudy the material), thereby demonstrating more sophisticated metacognitive knowledge regarding these specific topics.

The fact that this pattern did not hold for *Scenario 5: Spacing vs. massing* was unexpected. None of the groups (except the Seminar) endorsed the spacing outcome; in fact, they strongly endorsed massed study as the superior method, which parallels the metacognitive findings of Kornell and Bjork (2008). Why did students not realize the benefits of spaced study in this scenario? I predict that had I chosen a scenario describing more typical spacing versus massing situation (e.g., a student studying one hour per day over the course of seven days, versus seven hours in one day before an exam), participants who had learned about the spacing effect would have shown stronger endorsement of spacing. The fact that the scenario used in the current study instead described a spacing situation unfamiliar to the common use of the concept (i.e., the presentation of paintings in an interleaved or blocked fashion) may have driven the lack of metacognitive awareness that paintings in the interleaved (spaced) group would be better remembered. Students were unable to extrapolate their knowledge of the spacing effect to a novel situation presented on a far shorter time course and using a variety of exemplars within a given category (as opposed to strict repetition of the same material in a spaced versus massed fashion). Hence, the conclusion regarding lack of awareness of the spacing effect may be better couched as a lack of knowledge regarding the memorial benefits of interleaving over blocking of study materials (e.g., Richland et al., 2005). However, as expected, Seminar students who had read this original article were quite accurate in predicting the outcome.

Results from the global scenario performance variable (i.e., percentage of empirically-supported predictions) showed a clear pattern: Study 1 participants and those in Study 2 who did not receive targeted instruction were similar, and relatively low. Introductory psychology and Cognition course students who learned about applied memory topics were similar to each other, and outperformed the two 'non-instructed' groups. Finally, students in the Seminar group who read and discussed the specific outcomes presented in

---

[2] Given that cross-study comparisons should be interpreted with caution, it is important to note that even without the Study 1 participants in the analyses, patterns of contrasts were similar with regard to comparing the instructed groups to the non-instructed Introduction to Psychology group in Study 2.

Fig. 3 Global scenario performance, averaged across the six learning scenarios, in Studies 1 and 2: Percentage of participants indicating endorsement (i.e., rating ≥5) of the empirically-supported outcome. IP = *Introduction to Psychology*. Bars represent standard errors
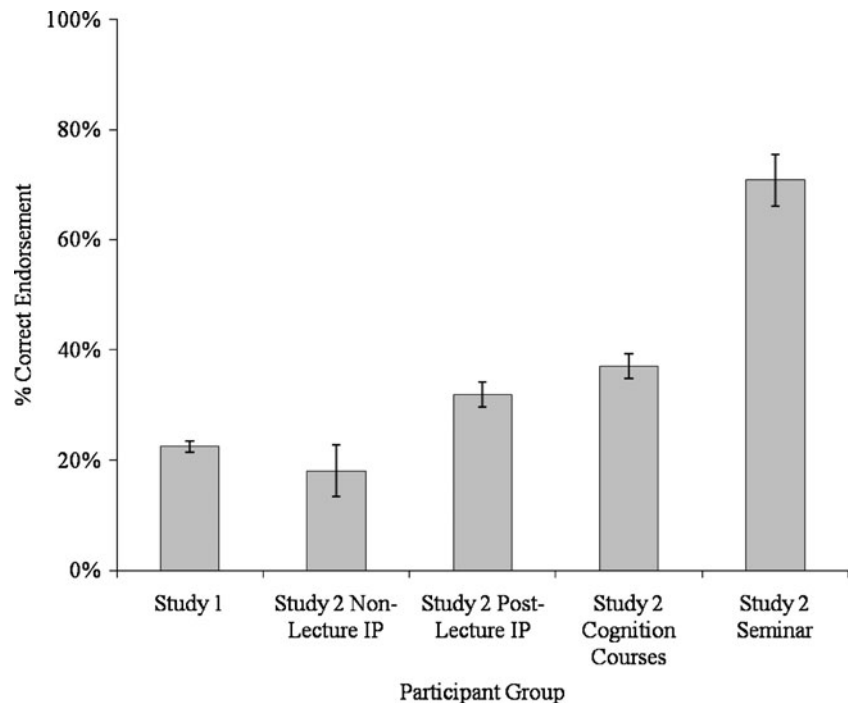
the survey had the highest performance compared to all groups. Notably, even though the Seminar group students technically 'knew' all the correct outcomes, performance on this measure was still not at ceiling.

Predictions regarding the MSR variable were not borne out; MSR was not predictive of performance on the survey items, nor correlated with demographics.

General discussion

This Internet-based survey study examined the extent to which undergraduates could predict which educational scenarios aide learning and memory for course material, in reference to what has been found to be effective in published research studies. Study 1 examined metacognitive awareness of six learning strategies (i.e., *Dual-Code Presentations, Static Media, Low-Interest Extraneous Details, Testing, Spacing,* and *Generating*) in a large and diverse sample of undergraduates, as well as correlations of scenario performance with an independent measure of metacognitive self-regulation (MSR). Study 2 used the same survey, with the goal of comparing groups who had received different levels of targeted instruction on applied learning and memory topics.

In Study 1, in contrast to research findings (Kornell & Bjork, 2008; Mayer et al., 2005, 2008; Roediger & Karpicke, 2006b), participants overall predicted that animated media (compared to static illustrations), high-interest (compared to low-interest) extraneous details, restudying (compared to taking a recall test), and massing (compared

to spacing/interleaving) the study of to-be-learned material would result in higher test scores. The lack of metacognitive awareness of the latter two topics is consistent with prior research (e.g., Roediger & Karpicke, 2006b; Kornell & Bjork, 2008). The one strategy that was weakly endorsed was generating one's own study materials, a finding consistent with past metacognitive research on the generation effect (e.g., Begg et al., 1991).

A global measure of metacognitive accuracy from the learning scenario rating scales revealed very low performance, with even the most accurate students only correctly endorsing the empirically supported options on a maximum of four out of six learning scenarios, and with nearly 1/5 of participants performing at floor level (0%). These findings are broadly consistent with the literature portraying poor metacognition and non-effective study strategies in undergraduates (e.g., Kornell & Bjork, 2007; Karpicke et al., 2009). As discussed previously, however, it is possible that metacognitive knowledge was underestimated in the current study due to the choice of relatively non-intuitive learning strategies. Interestingly, and consistent with predictions, MSR (e.g., Duncan & McKeachie, 2005) was positively correlated with global scenario performance. As a component of self-regulated learning (e.g., Pintrich, 2000), MSR may be a variable of interest in understanding real-world metacognitive accuracy across domains; however, because this correlational finding was small in magnitude and failed to replicate in Study 2, further investigation is warranted.

After establishing students' relatively poor understanding of several factors underlying learning and memory in Study 1, Study 2 was undertaken to compare scenario

performance among students experiencing different levels of specific instruction on applied learning and memory topics relevant to three of the scenarios in the survey (i.e., those in the category of *desirable difficulties*, Bjork, 1994). Results suggested that for the scenarios relevant to two of the instructed memory principles (i.e., testing, generation), all three instruction groups showed relatively strong endorsement of the empirically-supported outcomes, and outperformed both the control group and Study 1 participants. In contrast, for the scenario topics only explicitly studied by the advanced Seminar group (i.e., dual-coding, static media, low-interest details), along with the topic of spacing (inter-pretion of this outcome described in Study 2 Discussion), all non-Seminar groups failed to endorse the empirically-supported outcome. Global scenario performance increased from a low for Study 1 and Study 2 control group participants, to a medium level for the two groups instructed on the three memory topics, to a high for the Seminar students. Though this was not a true experiment with random assignment to conditions, and as such there is always the possibility of subject-selection effects given that students chose which course(s) to enroll in, this is precisely the pattern expected if depth of instruction in these areas indeed contributes to metacognitive knowledge of learning strategies.

Taken together, these two studies suggest a lack of metacognitive awareness of several specific learning and memory strategies relevant to educational contexts; and further, that targeted instruction in applied memory topics is associated with improved ability to predict the outcomes of learning scenarios. Seminar students, who learned about these topics directly from the primary sources, were most successful in scenario predictions. Though interesting, the conclusion from the Seminar group may have less practical importance because it is unrealistic that large numbers of undergraduates would be exposed to original research articles on study strategies. Also, the Seminar students were not perfect in their scenario predictions, suggesting that even at this advanced level of training, there is room for improvement in awareness and application of these principles.

This research contributes to the broader literature on applications of memory theory to higher education by providing an account of education-related metacognitive judgments spanning several specific research areas and cognitive theories. Further, the study is unique in eliciting ratings based on extrinsic cues, as opposed to intrinsic and/or real-time mnemonic cues (Koriat, 1997). The data patterns presented here, therefore, provide a picture of metacognitive knowledge for these six learning strategies uncontaminated by the types of metacognitive illusions that can arise from direct exposure to the to-be-learned materials (e.g., Karpicke, 2009). Results also suggest MSR as a variable of interest in understanding variations in students' understanding and application of learning and memory strategies.

The ultimate goal, of course, is improved metacognition for students' own day-to-day academic endeavors and, building on that, measurable changes in their study-related behaviors and course performance (e.g., Fleming, 2002; Tuckman, 2003). The current research does not address these latter points, and does not suggest that participants necessarily implemented their metacognitive knowledge; yet given the necessity of *knowing* about memory strategies before choosing to *apply* them to one's own behavior, the current studies are important in suggesting that (1) students' awareness of the effectiveness of several such strategies is low or non-existent, and (2) educational intervention, in the form of targeted instruction on learning and memory topics, may have the potential to improve metacognitive awareness of factors associated with academic success.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861–876. doi:10.1002/acp.1391

Azevedo, R., & Cromley, J. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523–535. doi:10.1037/0022-0663.96.3.523

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19,* 487–497. http://mc.psychonomic-journals.org/

Benjamin, A. S., & Bird, R. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language, 55*, 126–137. doi:10.1016/j.jml.2006.02.003

Benjamin, A., Bjork, R., & Schwartz, B. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55–68. doi:10.1037/0096-3445.127.1.55

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT.

Bloom, C. M., & Lamkin, D. M. (2006). The Olympian struggle to remember the cranial nerves: Mnemonics and student success. *Teaching of Psychology, 33*(2), 128–129. http://teachpsych.org/top/index.php

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527. doi:10.1080/09541440701326097

DeWinstanley, P.A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32,* 945–955. http://mc.psychonomic-journals.org/

Duncan, T., & McKeachie, W. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist, 40*(2), 117–128. doi:10.1207/s15326985ep4002_6

Dunlosky, J., & Nelson, T. (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. doi:10.1006/jmla.1994.1026

Fleming, V. (2002). Improving students' exam performance by introducing study strategies and goal setting. *Teaching of Psychology, 29*, 115–119. doi:10.1207/S15328023TOP2902_07

Hacker, D., Bol, L., Horgan, D., & Rakow, E. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170. doi:10.1037/0022-0663.92.1.160

Hartlep, K., & Forsyth, G. (2000). The effect of self-reference on learning and retention. *Teaching of Psychology, 27*, 269–271. doi:10.1207/S15328023TOP2704_05

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371. doi:10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469–486. doi:10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*, 471–479. doi:10.1080/09658210802647009

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. doi:10.1037/0096-3445.126.4.349

Koriat, A., & Bjork, R. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*(5), 1133–1145. doi:10.1037/0278-7393.32.5.1133

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. http://pbr.psychonomic-journals.org/

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592. doi:10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A., Eich, T., & Bjork, R. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25*, 498–503. doi:10.1037/a0017807

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*, 493–501. doi:10.1080/09658210902832915

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. doi:10.1037/0022-3514.77.6.1121

Maki, R., Foley, J., Kajer, W., Thompson, R., & Willert, M. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16*, 609–616. doi:10.1037/0278-7393.16.4.609

Mazzoni, G., & Nelson, T. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 21*(5), 1263–1274. doi:10.1037/0278-7393.21.5.1263

Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology. Applied, 14*, 329–339. doi:10.1037/a0013835

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology. Applied, 11*, 256–265. doi:10.1037/1076-898X.11.4.256

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52. doi:10.1207/S15326985EP3801_6

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320. doi:10.1037/0022-0663.90.2.312

Mayer, R. E., & Moreno, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368. doi:10.1037/0022-0663.91.2.358

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 481–493. doi:10.1080/09541440701326154

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4. doi:10.1207/S15326985EP3801_1

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., and Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. http://ncer.ed.gov

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). San Diego: Academic.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI.

Pyc, M. A., & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition, 38*, 431–440. doi:10.3758/MC.38.4.431

Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: generation and interleaving effects. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual conference of the cognitive science society*. Mahwah: Lawrence Erlbaum.

Roediger, H. L. I. I. I., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*, 183–186. doi:10.1111/j.1467-8721.2007.00500.x

Son, L. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 36*, 255–262. doi:10.1037/a0017892

Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30*, 601–604. doi:10.1037/0278-7393.30.3.601

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of*

Experimental Psychology. Learning, Memory, and Cognition, 18, 1074–1083. doi:10.1037/0278-7393.18.5.1074

Son, L. K., & Kornell, N. (2009). Simultaneous decisions at study: Time allocation, ordering, and spacing. *Metacognition and Learning, 4*, 237–248. doi:10.1007/s11409-009-9049-1

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research and Evaluation, 10*, 117–139. doi:10.1076/edre.10.2.117.27905

Tuckman, B. (2003). The effect of learning and motivation strategies training on college students' achievement. *Journal of College Student Development, 44*, 430–437. doi:10.1353/csd.2003.0034

University of Memphis Department of Psychology. (2008). *Lifelong learning at work and at home. 25 learning principles to guide pedagogy and the design of learning environments.* http://psyc.memphis.edu/learning/whatweknow/index.shtml

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*, 551–572. doi:10.1016/S0361-476X(02)00006-1

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44. http://pbr.psychonomic-journals.org/