Journal of Experimental Psychology: Applied
2011, Vol. 17, No. 3, 288–302

1076-898X/11/$12.00 DOI: 10.1037/a0024749

Accurate Monitoring Leads to Effective Control and Greater Learning of Patient Education Materials

Katherine A. Rawson, Rochelle O'Neil, and John Dunlosky
Kent State University

Effective management of chronic diseases (e.g., diabetes) can depend on the extent to which patients can learn and remember disease-relevant information. In two experiments, we explored a technique motivated by theories of self-regulated learning for improving people's learning of information relevant to managing a chronic disease. Materials were passages from patient education booklets on diabetes from NIDDK. Session 1 included an initial study trial, Session 2 included self-regulated restudy, and Session 3 included a final memory test. The key manipulation concerned the kind of support provided for self-regulated learning during Session 2. In Experiment 1, participants either were prompted to self-test and then evaluate their learning before selecting passages to restudy, were shown the prompt questions but did not overtly self-test or evaluate learning prior to selecting passages, or were not shown any prompts and were simply given the menu for selecting passages to restudy. Participants who self-tested and evaluated learning during Session 2 had a small but significant advantage over the other groups on the final test. Secondary analyses provided evidence that the performance advantage may have been modest because of inaccurate monitoring. Experiment 2 included a group who also self-tested but who evaluated their learning using idea-unit judgments (i.e., by checking their responses against a list of key ideas from the correct response). Participants who self-tested and made idea-unit judgments exhibited a sizable advantage on final test performance. Secondary analyses indicated that the performance advantage was attributable in part to more accurate monitoring and more effective self-regulated learning. An important practical implication is that learning of patient education materials can be enhanced by including appropriate support for learners' self-regulatory processes.

Keywords: self-regulated learning, patient education, metacognative monitoring, idea-unit judgments, self-testing

Patients' learning of medical information is associated with a variety of health related outcomes, including improved self-management skills, preventive care, adherence to prescribed regimens, and reduced health care costs, hospitalizations, and morbidity (e.g., Barat, Andreasen, & Damsgaard, 2001; Gordon, Street, Kelly, Souchek, & Wray, 2005; Jacobson et al., 1999; Williams, Baker, Honig, Lee, & Nowlan, 1998). Learning medical information is particularly crucial for managing some chronic diseases. For example, individuals who are diagnosed with diabetes must learn how to prevent and detect the numerous problems that can result from the disease. If not properly managed, diabetes can lead to blindness, kidney failure, amputations, and cardiovascular disease. Poor comprehension and memory for relevant medical information is a health risk in itself, because not fully understanding or remembering essential information (e.g., how to check blood-sugar levels, knowing the early warning signs of retinal damage) may increase the risk of debilitating or life-threatening complications. Accordingly, our primary goal was to explore a new technique—inspired by theories of self-regulated learning—to improve people's learning of materials relevant to managing a chronic disease (henceforth, called *patient education materials*). Before we describe the current approach, we first discuss the importance of patient education materials and the modal approach for improving people's learning of them.

The primary responsibility for educating individuals about disease-relevant information is increasingly shifting from practitioners to patients themselves. Although physicians explain some information to patients during appointments, patients often do not ask questions to clarify potential misunderstandings (Cegala, 1997). Doctors often underestimate how much information patients want, and patients are often dissatisfied with the amount of information they receive (Kenny et al., 1998; Ong, de Haes, Hoos, & Lammes, 1995; Suhonen & Leino-Kilpi, 2006). Gazmararian, Williams, Peel, and Baker (2003) note that this trend is attributable in no small part to "the pressures of today's health care environment; physicians and nurses have less time for patient education. To deal with the lack of time to instruct patients in self-management, doctors and nurses often rely on the written word" (p. 274). Even when information is provided verbally during consultation, patients often report still wanting additional written information (Sheard & Garrud, 2006).

Katherine A. Rawson, Rochelle O'Neil, and John Dunlosky, Department of Psychology, Kent State University.

The research reported here was supported by a James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind, and Behavior Collaborative Award.

Correspondence concerning this article should be addressed to Katherine A. Rawson, Kent State University, Department of Psychology, P.O. Box 5190, Kent, OH 44242-0001. E-mail: krawson1@kent.edu

Given that patients are increasingly provided with supplementary materials that they are responsible for learning, an important question arises: How should these patient education materials be presented to increase comprehension and retention of the most critical information? One approach to answering this question focuses on improving the patient education materials themselves, which are often poorly written for their purpose (Gal & Prigat, 2005; Perkins & Cohen, 2008; Wilson & Wolf, 2009). Part of the problem lies in the mismatch between the reading level of patients and the readability of the materials. The 2003 National Assessment of Adult Literacy estimated that 93 million Americans have only a basic or below basic level of prose literacy (White & Dillow, 2005), and various studies have reported average adult reading levels at or below an eighth grade level (e.g., Cotugna, Vickery, & Carpenter-Haefele, 2005; Davis, Crouch, Wills, Miller, & Abdehou, 1990; Hoffmann & McKenna, 2006; Winslow, 1998). Unfortunately, patient education materials often require much higher reading ability (e.g., Hoffmann & McKenna, 2006; Singh, 2000; Wolf, Davis, Shrank, Neuberger, & Parker, 2006). For instance, Perkins and Cohen (2008) estimated the readability of 235 written patient education materials across a wide variety of specialties and found a mean readability level of Grade 10, with some materials reaching a reading level of Grade 15.

Given such evidence, the modal recommendation for enhancing patient education has been to improve the readability of these materials (e.g., Baker, Newton, & Bergstresser, 1988; Cotugna et al., 2005; Davis et al., 1990; Singh, 2000; Vallance, Taylor, & Lavallee, 2008). Revisions have included substituting simpler terms for words with multiple syllables, defining new or difficult words (i.e., medical terms), and decreasing sentence length. Improving readability can enhance learning (e.g., Baker et al., 1988; Dowe, Lawrence, Carlson, & Keyserling, 1997; Estey, Musseau, & Keehn, 1994), although the gains are often modest. Other recommendations for modification have included using simple fonts, using list formats when possible, and using positive wording and avoiding negations (e.g., Wilson & Wolf, 2009). Importantly, in the vast majority of prior studies, changes to improve patient learning have been made only to the linguistic characteristics of the content or physical appearance of the materials. Although these changes can be beneficial, our proposal is that learning can be further improved by providing external supports that scaffold and in turn improve the effectiveness of people's self-regulated learning. Effective self-regulated learning involves the interplay among cognitive and metacognitive processes, such as goal setting, accurate monitoring of learning, and control of study (Dunlosky & Ariel, 2011; Pintrich, 2004; Winne & Hadwin, 1998; Zimmerman, 1990).

Most important for the current research, our approach provides external supports for improving people's monitoring and control while they learn patient education materials. *Monitoring* refers to people's evaluations of their own learning, and *control* involves making study decisions, such as deciding what to study, how long to study, and what study strategies to use. A central assumption of general theories of self-regulated learning is that accurate monitoring is crucial for effective control, because people use their monitoring to inform their control decisions (e.g., deciding what to study and how long to study based on their evaluation of how well they have learned the material). For example, if a patient believes that he will not forget which symptoms indicate the warning signs of retinal damage, he may decide not to study those symptoms further; of course, if his judgment is incorrect and in fact he does forget some of the symptoms, then such inaccurate monitoring will lead to ineffective control. By contrast, if the patient judged that he had not learned the symptoms well enough to remember them later, he could decide to restudy that information at another time. In this manner, accurate monitoring of learning is expected to support more effective control decisions, which in turn supports better learning and retention (Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Lipko et al., 2009; Winne & Hadwin, 1998).

This assumption has been recently validated in two studies by Thiede and his colleagues (Thiede, 1999; Thiede, Anderson, & Therriault, 2003). For example, Thiede (1999) had people study Swahili-English word pairs during an initial study trial. They then made a judgment of learning (JOL) for each item, in which they predicted the likelihood of later remembering the English response when shown the Swahili cue. JOLs were negatively correlated with item selection, such that participants were more likely to restudy items given a low JOL—in other words, participants more often chose to restudy items that they predicted they would not be able to remember later (or that were more difficult to learn, Son & Metcalfe, 2000). Most important, participants who made the most accurate JOLs demonstrated higher final test performance, presumably because the higher levels of judgment accuracy led to better study decisions (for converging evidence involving text materials, see Thiede et al., 2003).

Based on this evidence and general theories of self-regulated learning (Dunlosky & Ariel, 2011; Winne & Hadwin, 1998), the current experiments were designed to improve people's learning of patient education materials by providing external support that encouraged learners to explicitly and accurately monitor their ongoing learning. A critical aspect of this intervention involves prompting people to test themselves on the to-be-learned concepts. For instance, at some point after studying key information about diabetes (e.g., the symptoms associated with retinal damage), people are prompted to recall the relevant information. In this case, they would be prompted with "What are the symptoms of diabetes-related eye problems?" and asked to type a response. These delayed self-tests are expected to improve learning in two ways. First, testing can promote accurate monitoring (for reviews, see Dunlosky, Rawson, & McDonald, 2002; Rhodes & Tauber, in press). For example, when people are unable to recall anything when prompted, the failed recall attempt provides evidence that the material is not sufficiently learned and thus should be restudied. Of course, if they correctly recall the information, doing so would indicate that less time is needed for restudying that information. Second, correctly recalling a concept provides an additional benefit, because doing so can boost memory for the focal information (for recent reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006).

In summary, self-testing has dual benefits: it can improve people's monitoring accuracy and hence the effectiveness of their control of learning, and it can directly boost memory for those concepts that are correctly recalled. Accordingly, we evaluated the efficacy of incorporating support for self-testing and monitoring into patient education materials. Our base materials included passages from patient education booklets from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). All participants completed three sessions. In the first session, each of

the passages in the base materials was presented for one experimenter-paced study trial that was designed to represent a patient education session in which a health educator leads a presentation of information to a group of patients.

The second session was designed to represent the situation in which individuals are given patient education materials for further study on their own. Accordingly, in Session 2, participants regulated their own learning of the base material for 50 minutes. Learners in the *unguided restudy* group were presented with a menu of all the information topics on the main screen and were permitted to use the time however they wished to study the material. This presentation format represents a business-as-usual comparison group, because people typically study patient education materials with minimal guidance. In the *self-regulated learning support* group (henceforth, the *SRL support group*), on each trial, learners were presented with a prompt to test themselves on key information (e.g., "What can you do to prevent diabetes eye problems?"). They then evaluated the quality of their recall response by rating it from 0 (I could not remember any of the relevant information) to 100 (I remembered all of the relevant information accurately). After making this *self-score judgment*, they were given the option to restudy relevant information before moving on to the next prompt. The SRL support group thus differed from the unguided restudy group in several respects; not only did learners in this group overtly self-test and make monitoring judgments, they were also shown prompts that they would be tested on during the final recall test (i.e., the prompts used to elicit self-tests during learning were the same prompts used on the final recall test). Although participants in the SRL support group were not told that the prompts on the final test would be the same as those presented during practice, the prompts alone may improve performance by signaling the key information. To evaluate the extent to which any performance advantage for the SRL support group was attributable to the presentation of the prompts per se (rather than to the self-testing and monitoring afforded by the prompts), the *prompts only group* was shown the same prompts and restudy menus on each trial as in the SRL support group, but without overt self-testing and monitoring. In the third session, all participants completed a final recall test over the material.

Based on theory of self-regulated learning described earlier, our focal prediction is that final test performance will be significantly greater for the SRL support group than for the unguided restudy group and the prompts only group. To revisit, this prediction is based on the assumptions that accurate monitoring leads to better control and restudy decisions, which in turn enhance learning (as explained later, we also conducted secondary analyses to test these assumptions). To foreshadow, results from Experiment 1 demonstrated that the SRL support group significantly outperformed the other two groups on the final test. However, performance was relatively modest, so Experiment 2 was designed to improve the efficacy of the SRL intervention by providing further support to enhance the accuracy of people's monitoring judgments.

Experiment 1

Method

Participants and design. Participants included 81 undergraduates from a large Midwestern university who participated for course credit. The sample was 65% female, 80% Caucasian, 14% African American, 4% Latino, 2% American Indian/Alaska Native, and 2% Asian (participants were able to report more than one ethnicity), with a mean age of 19 (range 17–39). For demographic purposes, participants completed two health literacy measures, the short form of the Test of Functional Health Literacy in Adults (sTOFHLA, Baker, Williams, Parker, Gazmararian, & Nurss, 1999) and the Medical Term Recognition Test (METER, Rawson et al., 2009). Mean sTOFHLA score was 34.8 of 36 ($SE = 0.2$, range 21–36) and mean METER score was 37.1 of 40 ($SE = 0.2$, range 28–40), indicating that the majority of the sample had adequate health literacy.

Participants were randomly assigned to one of three groups (unguided restudy group, SRL support group, or prompts only group).

Materials. The materials consisted of 35 passages from seven booklets from the NIDDK *Prevent Diabetes Problems* series (*Keep Your Diabetes under Control, Keep Your Eyes Healthy, Keep Your Feet and Skin Healthy, Keep Your Kidneys Healthy, Keep Your Heart and Blood Vessels Healthy, Keep Your Nervous System Healthy,* and *Keep Your Teeth and Gums Healthy*). The 35 passages included two introductory passages providing an overview of diabetes and 33 target passages grouped by six main topics, including eyes, feet and skin, heart and blood vessels, kidneys, nervous system, and teeth and gums. Each main topic included three to five passages ranging from a total of 519–1202 words and including 0–3 figures. Individual passages ranged from 54–365 words (see Figure 1 for a sample passage). Two questions were developed for each of the six main topics, with the format "What can you do to prevent diabetes problems with _____?" and "What are the symptoms of diabetes problems with _____?". These 12 questions were used as prompts during Session 2 for the SRL support group and prompts only group (described below) and were also the final test questions used in Session 3. All instructions and materials were presented via computer.

Procedure. In Session 1, each individual completed an informed consent form and was then seated at a computer carrel. The computer first presented background information about diabetes and general instructions about the materials participants would be learning. Participants were then informed of the purpose of the study ("to explore how to help people with diabetes better understand and remember the important information they need to know to effectively manage their disease") and were given an overview of each of the three experimental sessions (experimenter-paced initial study during Session 1, self-paced restudy during Session 2, and a final test during Session 3). Participants subsequently received detailed instructions for the tasks they would be completing during Session 1 (participants were not informed of their particular group assignment prior to Session 2). They were asked to act as though they had been recently diagnosed with diabetes and were attending a patient education program to learn about the disease and were encouraged to take the experiment as seriously as possible.

After reading the instructions, participants then completed a series of demographic questions relating to diabetes including: *Do you have diabetes?* (of the 77 who responded, four participants responded yes); *If yes, how old were you when you were diagnosed?* (M = 19 years), *If yes, how old are you now?* (M = 20 years); *If yes, have you attended a patient education program in*

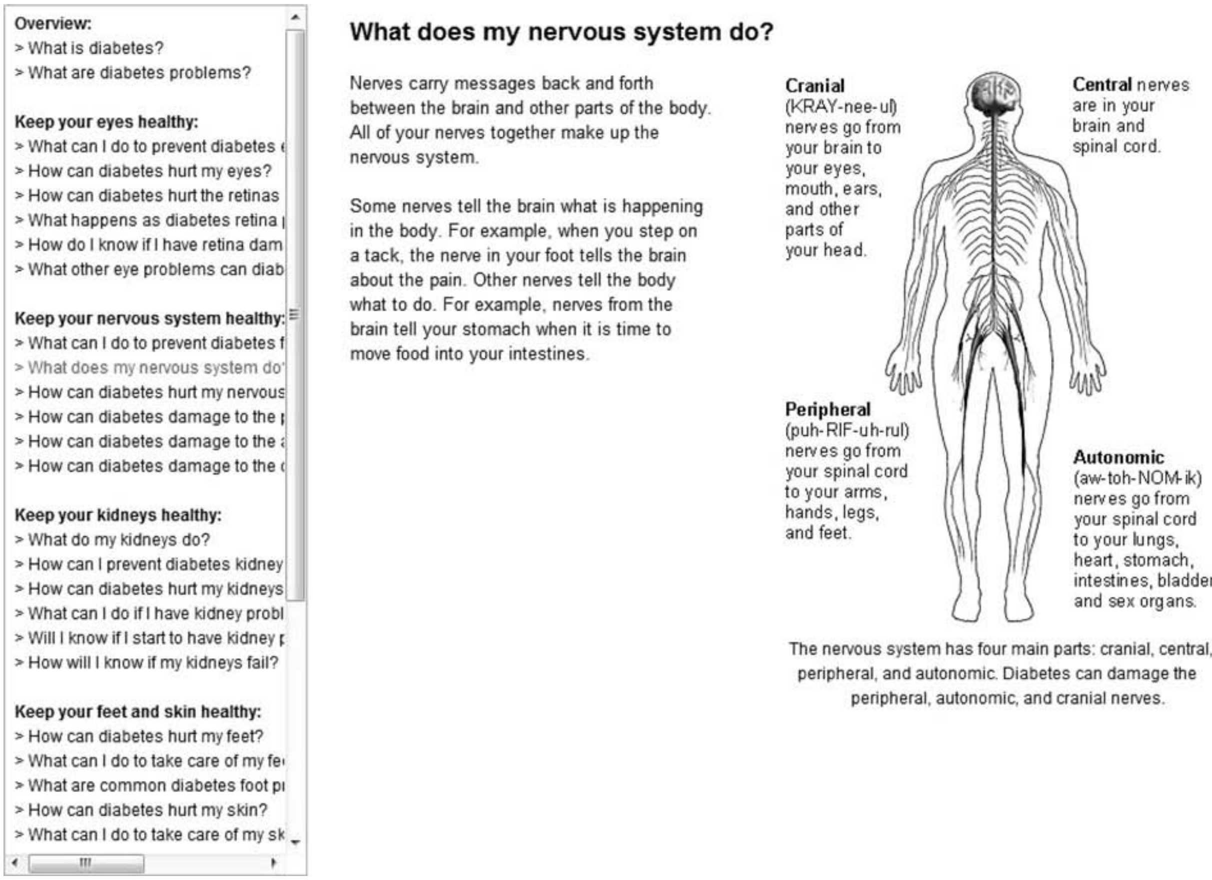

Figure 1. Illustration of the screen layout used to present passages during Session 1. Screen layout for presentation of passages selected for restudy during Session 2 was the same, except that a button labeled "Done Studying" was also presented at the bottom of the screen.

the past? (0% responded yes), and *Do you know anyone who has diabetes?* (72% responded that they had at least one friend or family member with diabetes; of those who responded yes, the average number of people known with diabetes was 2, range 1–6). Next, participants completed a prior knowledge questionnaire consisting of eight short-answer questions (e.g., "What is diabetes?").

Participants were then presented with the two introductory passages and each of the 33 target passages one at a time for an initial study trial. Each passage was presented via computer at a predetermined rate based on word length and whether it contained a figure (350 milliseconds per word plus an additional 15 seconds for each figure). On each screen, participants were notified when they had 15 seconds remaining to study the passage. As illustrated in Figure 1, each passage was presented in the main portion of the screen, and the left side of the screen contained a menu vertically listing the topics that were to be presented over the course of the session; the current passage was marked with red font as they progressed through the program. After initial presentation of the material was complete, participants were reminded to return in two days for Session 2.

Session 2 consisted of a 50-min study period. All participants were first reminded that they should act as though they had been diagnosed with diabetes and that they would be tested on their learning of the information during Session 3. Participants then received task instructions specific to their assigned group. The unguided restudy group was then presented with a menu of 33 passages, categorized by the six main topics (see Figure 2). Participants could choose to study a passage by clicking on the button next to that subtopic. The information appeared in the main window (similar to the one displayed in Figure 1) and stayed visible until the participant clicked on a "done studying" button presented at the bottom of the screen to return to the main menu of passages. Participants were permitted to use the 50 minutes however they wished to study the passages.

For the SRL-support group, each trial began with the presentation of a prompt (e.g., "What can you do to prevent diabetes eye problems?") along with a blank text field and instructions to type in a response. Participants indicated that their response was complete by clicking on a "Done" button at the bottom of the screen. After participants submitted a response, they were asked to evaluate the quality of their response. The response was presented along with the prompt, and participants were asked, "How well do you think you have answered this question?" Participants made this self-score judgment using a slider scale from 0 (I could not remember any of the relevant information) to 100 (I remembered all of the relevant information accurately). After making their

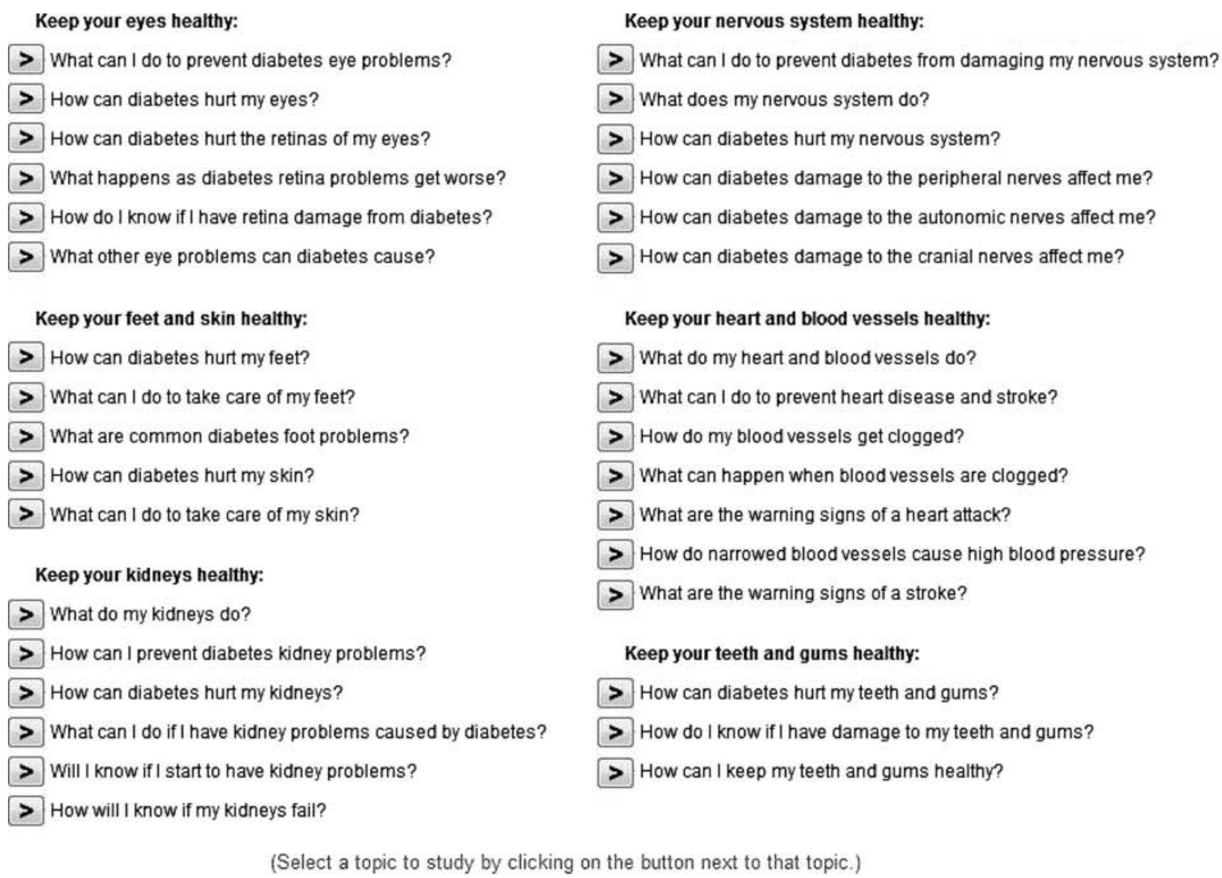

Figure 2. Illustration of the screen layout used to present the main menu to participants in the unguided restudy group during Session 2.

judgment, their recall response remained on the screen and participants were also then shown a list of the passages relevant to that topic (see Figure 3). Participants were free to choose to restudy one or more of the relevant passages from the list by clicking on the corresponding button, or they could choose to move on to the next prompt without any further study. The 12 prompts were repeated in the same order until the 50-min study period was complete.

Finally, for the prompts only group, on each trial participants were shown a prompt along with the list of relevant passages but *without* the overt recall and self-score judgments (similar to the screen displayed in Figure 3, except without the response field). Participants could choose to restudy one or more of the relevant passages from the list, or they could choose to move on to the next prompt. The 12 prompts were repeated in order until the 50-min study period was complete.

Session 3 took place five days after Session 2 and involved a final recall test. Participants were instructed that they would be tested to evaluate how much information they could remember from the previous two sessions. They were instructed to do their best to answer each prompt as completely and accurately as possible. Each of the 12 prompts was presented one at a time on the screen, and participants were asked to recall the relevant information.

Results and Discussion

Prior knowledge about diabetes. Before describing final test performance, we first consider prior knowledge responses to confirm that the groups had similar levels of preexperimental knowledge about diabetes. Prior knowledge responses were scored as total number of words generated. Overall, participants' responses were relatively brief (words per question: $M = 10.1$, $SE = 1.6$, in the unguided restudy group, $M = 8.0$, $SE = 1.1$, in the SRL support group; and $M = 12.5$, $SE = 2.4$, in the prompts only group), and the three groups did not differ significantly, $F(2, 75) = 1.46$.

Final test performance. For the recall questions in Session 3, target responses included any information explicitly stated in one or more of the 33 target passages that provided a direct answer to that specific question. For example, for the question, "What can you do to prevent diabetes eye problems?", target responses included "Keep blood sugar/glucose level normal, Keep blood pressure normal, Have eyes examined each year, If pregnant, see eye care professional in first three months, and Don't smoke." Across all 12 questions, there were 124 total target idea units, ranging from 4–26 per question ($M = 10.33$). For each response to a question, we scored the number of target ideas correctly recalled for that question (including verbatim statements and correct para-

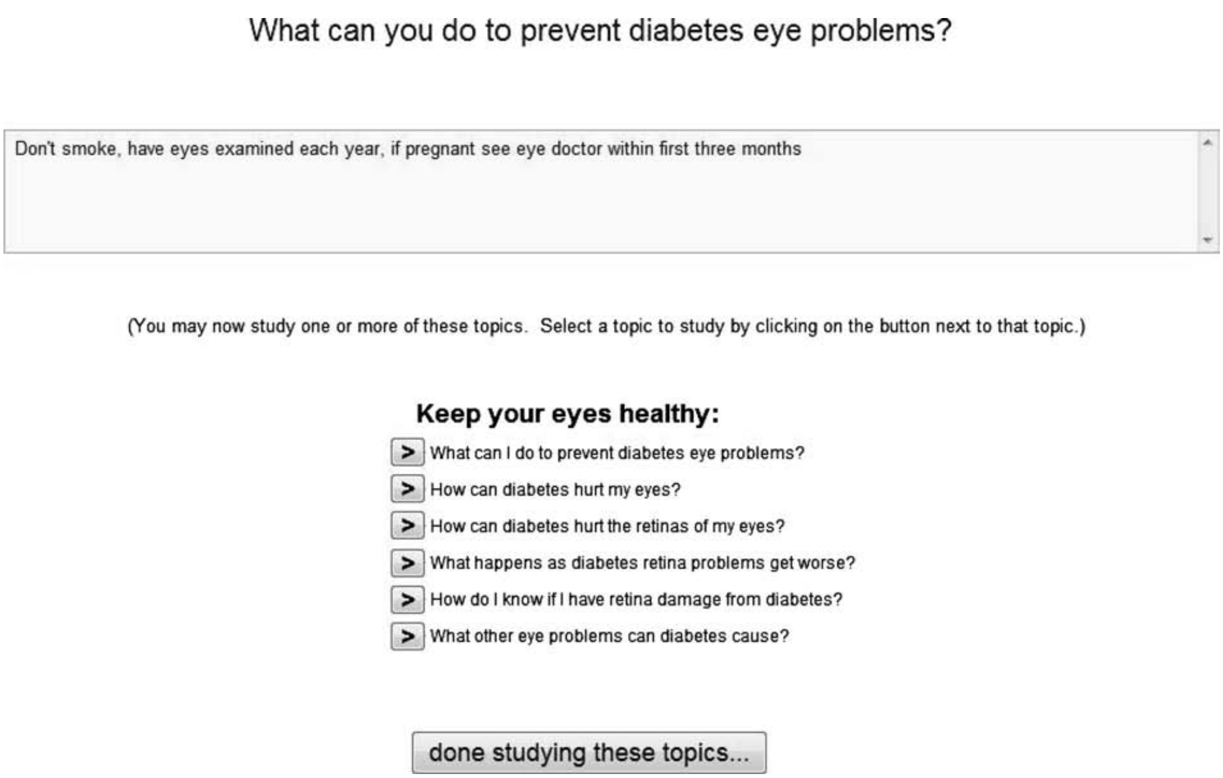

Figure 3. Illustration of the screen layout used to present the restudy menu to participants in the SRL support group during Session 2. A prompt is shown at the top of the screen, along with the participant's response in an uneditable text field.

phrases). Concerning the grain size of scoring, an idea unit was generally equivalent to a simple noun-verb phrase.

For each participant, we computed the total number of idea units correctly recalled across the 12 final test questions. Mean final test performance for each group is presented in Figure 4. A one-way analysis of variance (ANOVA) revealed a significant effect of group, $F(2, 76) = 5.59$, $MSE = 61.45$, $p = .005$, $\eta_p^2 = .13$. Follow-up *t* tests showed that the SRL support group significantly outperformed the prompts only group and the unguided restudy group [$t(49) = 3.12$, $p = .003$, $d = 0.87$, and $t(50) = 2.30$, $p = .026$, $d = 0.64$, respectively]. Performance in the prompts only group and the unguided restudy group did not significantly differ, $t(53) = .84$, $d = 0.23$, suggesting that the advantage of the SRL support group over the other two groups was not due to the presentation of prompts per se.

Although we had no a priori prediction about whether the prompts-only group would outperform the unguided restudy group, this null effect is perhaps surprising given the potential for transfer-appropriate processing; that is, the prompts presented during practice and for the final test were identical, and this nominal match may produce functionally similar processing during study and later retrieval. In particular, for both cases the prompt may trigger a search of long-term memory for the sought-after information. Based on this reasoning, one explanation is that participants in the prompts-only group did not spontaneously use the prompts to search memory for the relevant information during learning. For instance, they may have merely read the prompt without an exhaustive search of memory (as in Dunlosky et al., 2005). If so, then the processing triggered by the prompt at study would be qualitatively different than the processing triggered by the prompt during the final test. This possibility could be evaluated in future research, but we do not discuss it further here, given that our focus was on how SRL support could enhance learning.

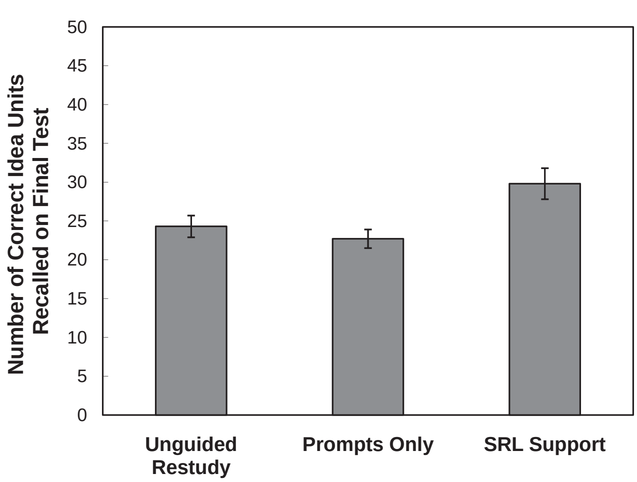

Figure 4. Mean number of idea units correctly recalled on the final test as a function of group in Experiment 1. Error bars report standard errors of the means.

As predicted, the SRL-support group did significantly outperform the other two groups, but the effect was relatively modest (only a 21% relative improvement over the other two groups). Furthermore, the absolute level of performance achieved was low, with only 24% of the total possible idea units correctly retrieved. Why were the effects of self-regulated learning support so modest? To explore potential answers to this question, we next examine the efficacy of monitoring and control during Session 2.

Relative accuracy of judgments. Concerning monitoring accuracy, we examined both relative accuracy and absolute accuracy of the self-score judgments made during Session 2 for the SRL support group. For each individual, we operationalized relative accuracy as the Goodman-Kruskal gamma correlation between the recall score on self-testing trials and the corresponding self-score judgments across all trials. The mean across individual correlations is reported in Table 1. Relative accuracy was significantly greater than zero, $t(22) = 8.63$, $p < .001$, indicating that learners tended to assign higher self-scores to responses in which they correctly recalled more information. However, relative accuracy was only moderate at best.

Absolute accuracy of judgments. As further evidence for limited monitoring accuracy, we examined one aspect of absolute accuracy, namely, the degree to which participants were overconfident in their commission errors. Commission errors are common and refer to cases in which a participant generates a response that is entirely incorrect (as opposed to omission errors, in which no response is generated). Note that any self-score judgment above 0 for a commission error represents overconfidence, and overconfidence here can undermine effective control (e.g., if individuals choose not to study information they falsely believe they already know). To evaluate overconfidence in commission errors, participants' recall responses from Session 2 were separated into four response categories: responses in which 1–2 idea units were correctly recalled, responses in which 3–4 idea units were correctly recalled, responses in which five or more idea units were recalled, and trials in which participants gave a response but none of the information contained in the response was correct (commissions). Thus, any response including a mix of correct and incorrect information was not counted as a commission but instead was included in one of the other three categories. Omissions were rare, and therefore data were insufficient for analysis of this category. For each individual, we then computed the mean self-score judgment assigned to responses in each category. Means across individuals are reported in Figure 5.

Self-score judgments were significantly lower for responses with 1–2 correctly recalled idea units versus 3–4 correctly recalled idea units [$t(22) = 6.51$, $p < .001$, $d = 0.85$], which in turn were lower than for responses with 5 or more correctly recalled idea units [$t(16) = 3.15$, $p = .006$, $d = 0.40$].[1] This increasing trend—responses with more correct idea units recalled judged as more complete and accurate—underlies the above-chance relative accuracy discussed previously. Most important, participants were overconfident at judging commission errors. Self-score judgments for commissions were significantly greater than zero [$t(21) = 12.65$, $p < .001$], and they did not significantly differ from self-score judgments for responses with 1–2 correctly recalled idea units [$t(21) = 1.06$, $p = .303$, $d = 0.16$].[2] This overconfidence is a fundamental problem and suggests that participants are not entirely accurate in identifying when they incorrectly recall target information for a given prompt. If these self-score judgments in turn influence control decisions, restudy may be suboptimal. Thus, we turn to the question: To what extent were self-score judgments related to control decisions?

Control of restudy during Session 2. Before we examine the relationship between monitoring and control in the SRL support group, we provide a brief summary of restudy measures in all three groups during Session 2. For each participant, we calculated how many of the 33 target passages were restudied at least once. A one-way ANOVA revealed a significant effect of group, $F(2, 74) = 12.62$, $MSE = 11.34$, $p < .001$, $\eta_p^2 = .25$. More passages were restudied at least once in the unguided restudy group ($M = 31.8$, $SE = 0.5$) and in the prompts only group ($M = 30.4$, $SE = 0.6$) than in the SRL support group ($M = 27.1$, $SE = 0.9$), $t(48) = 4.57$, $p < .001$, $d = 1.29$, and $t(49) = 3.17$, $p = .003$, $d = 0.89$, respectively. Given that participants were free to restudy a passage more than once, we also calculated the total number of times passages were

Table 1
Experiments 1 and 2: Relative Accuracy of Monitoring Judgments

	M	*SE*
Experiment 1		
SRL-support group	.33*	.04
Experiment 2		
SRL-support group	.44*	.04
IDU group–self-score judgments	.45*	.05
IDU group–idea unit judgments	.71*	.05

Note. Values represent the mean across individual gamma correlations between judgments (self-score or idea unit judgments) and the number of idea units correctly recalled across trials during Session 2. M = mean; SE = standard error.
* Correlation significantly greater than zero, $p < .05$.

[1] Responses were scored with respect to the number of correct ideas, whereas participants made self-score judgments using a continuous scale that did not require participants to quantify the number of ideas they believed they had recalled. Thus, judgments for responses containing one or more correct ideas cannot be meaningfully interpreted as overconfident or underconfident. In contrast, because responses categorized here as commission errors contained no correct information, a judgment greater than zero on any scale indicates overconfidence.

[2] This null effect may be consistent with the accessibility hypothesis (Koriat, 1993), which is that people base their judgments on the amount of information accessed instead of an evaluation of the quality of what is accessed (but see Dunlosky et al., 2005). Consistent with this possibility, participants recalled the same amount for commission errors and for responses containing 1–2 correct ideas. In Experiment 1, mean number of words per response was 16.0 ($SE = 2.1$) for commissions, 16.0 ($SE = 1.6$) for responses containing 1–2 correct ideas, 22.5 ($SE = 2.3$) for responses containing 3–4 correct ideas, and 27.7 ($SE = 3.3$) for responses containing 5+ correct ideas. In Experiment 2, mean number of words per response was 13.9 ($SE = 1.4$) for commissions, 14.7 ($SE = 0.8$) for responses containing 1–2 correct ideas, 18.9 ($SE = 0.7$) for responses containing 3–4 correct ideas, and 25.4 ($SE = 1.3$) for responses containing 5+ correct ideas. A major point here is that retrieving incorrect information can lead to a metacognitive illusion of knowing, but to foreshadow, this illusion was largely diminished by the use of idea-unit judgments in Experiment 2.

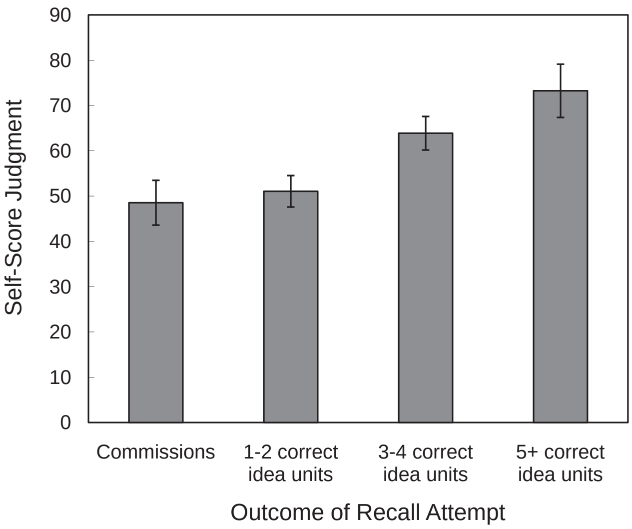

Figure 5. Mean self-score judgment as a function of the number of correct idea units actually contained in a self-test response during Session 2, for the SRL support group in Experiment 1. Error bars report standard errors of the means.

restudied (i.e., *restudy events*). The total number of restudy events was similar across groups (unguided restudy group, $M = 69.5$, $SE = 5.7$; prompts only group, $M = 66.0$, $SE = 5.2$; SRL support group, $M = 56.6$, $SE = 3.7$), $F(2, 74) = 1.73$, $MSE = 642.52$, $p = .185$, $\eta_p^2 = .05$. Finally, we also calculated the mean amount of restudy time (in seconds) spent per restudy event for each participant. A one-way ANOVA revealed a significant effect of group, $F(2, 74) = 19.82$, $MSE = 267.06$, $p < .001$, $\eta_p^2 = .35$. Time per restudy event was greater in the unguided restudy group ($M = 41.4$, $SE = 3.6$) and in the prompts only group ($M = 39.4$, $SE = 3.2$) than in the SRL support group ($M = 15.1$, $SE = 2.7$), $t(48) = 5.76$, $p < .001$, $d = 1.63$, and $t(49) = 5.70$, $p < .001$, $d = 1.60$, respectively. Thus, the advantage in final test performance for the SRL support group was not attributable to participants in this group restudying more passages or restudying them longer. If anything, the trends were consistently in the opposite direction.

Relationship between monitoring and control in the SRL support group. To evaluate the extent to which monitoring was related to control decisions, we examined two measures of control for the SRL support group during Session 2. First, for each individual, we computed the correlation across all trials between the self-score judgments and the total number of passages studied (i.e., selected from the topic menu during the subsequent restudy opportunity) after each prompt. The mean across individual correlations (see Table 2) was significantly less than zero, $t(23) = 4.29$, $p < .001$. Second, for each individual, we also computed the correlation across all trials between the self-score judgments and the total time spent studying these passages during the subsequent restudy opportunity following each prompt. The mean across individual correlations (see Table 2) was significantly less than zero, $t(23) = 5.83$, $p < .001$. Thus, as a learner's self-score judgment increased, the number of total passages studied decreased and the total time studying passages decreased. Results of these secondary analyses suggest that the restudy decisions of participants in the SRL support group were at least partially guided by their metacognitive judgments.

Relationship between control decisions and learning gains. To revisit, theory of self-regulated learning assumes that monitoring accuracy feeds into control decisions, and control decisions in turn are related to learning. Accordingly, we examined the relationship between control decisions during study and learning. Note that performance on the first test trial in Session 2 reflected the individual's level of knowledge (either from prior knowledge and/or from acquisition during Session 1) prior to self-regulated learning, and this baseline performance predicted final test performance (across participants, $r = .45$, $SE = .04$). Thus, because our interest was in the relationship between control decisions and learning, we used learning gain (rather than absolute level of final test performance) in the analyses below. *Learning gain* for each participant was computed as the difference score between final test performance and performance on the first test trial in Session 2 for each item. Larger scores indicate greater gains in performance.

For each individual, we computed the correlation between (a) the total number of passages studied during the restudy opportunities and (b) learning gain. The mean across individual correlations (see Table 3) was greater than zero, $t(23) = 3.85$, $p = .001$. We also computed the correlation between (a) the total time spent studying all passages during the restudy opportunities and (b) learning gain. As reported in Table 3, the mean across these individual correlations was also greater than zero, $t(23) = 5.02$, $p < .001$. Thus, results confirmed that control decisions during restudy were related to learning gains.

Experiment 2

Experiment 1 demonstrated that including support for self-regulated learning within patient education materials is beneficial for learning. However, the improvement in learning was relatively modest, and secondary analyses revealed that monitoring accuracy

Table 2
Experiments 1 and 2: Correlations Between Monitoring Judgments and Control Variables

	M	*SE*
Experiment 1		
SRL-support group, self-score judgments		
Judgment–number of passages studied	−.20*	.05
Judgment–total study time for all passages studied	−.21*	.04
Experiment 2		
SRL-support group, self-score judgments		
Judgment–number of passages studied	−.11*	.04
Judgment–total study time for all passages studied	−.10*	.04
IDU group, self-score judgments		
Judgment–number of passages studied	−.19*	.06
Judgment–total study time for all passages studied	−.19*	.05
IDU group, idea unit judgments		
Judgment–number of passages studied	−.38*	.06
Judgment–total study time for all passages studied	−.35*	.05

Note. Values represent the mean across individual gamma correlations between judgments (self-score or idea unit judgments) and two control measures, the number of passages studied and total study time, across trials during Session 2. M = mean; SE = standard error.
* Correlation significantly less than zero, $p < .05$.

Table 3
Experiments 1 and 2: Correlations Between Control Variables and Learning

	M	*SE*
Experiment 1		
SRL-support group		
Learning gain–number of passages studied	.20*	.05
Learning gain–total study time for all passages studied	.22*	.04
Experiment 2		
SRL-support group		
Learning gain–number of passages studied	.03	.07
Learning gain–total study time for all passages studied	.01	.07
IDU group		
Learning gain–number of passages studied	.22*	.05
Learning gain–total study time for all passages studied	.22*	.05

Note. Values represent the mean across individual gamma correlations between learning gain (the gain in the number of idea units correctly recalled from the first trial of Session 2 to the final test) and two control measures, the number of passages studied and total study time in Session 2. M = mean; SE = standard error.
* Correlation significantly greater than zero, $p < .05$.

in the SRL support group was moderate at best. If a person does not accurately monitor what he or she has learned, it will constrain effective control, which ultimately constrains learning. The goal of the current experiment was to improve monitoring with the expectation of improving control, and ultimately performance.

Fortunately, monitoring accuracy can be improved by providing learners with appropriate standards for comparison when evaluating response quality (Dunlosky, Hartwig, Rawson, & Lipko, 2011; Lipko et al., 2009; Rawson & Dunlosky, 2007). In Lipko et al. (2009), middle school students studied definitions about literary nonfiction. After initial study, the students recalled as much of each definition as they could and then made a self-score judgment for each response. When evaluating the quality of their responses, one group of students was not given any type of standard (as was the case for the SRL support group in the current Experiment 1). Another group was shown the correct definition broken down into its constituent idea units, and the students were asked to check off which of the ideas they had included in their response. Results showed that participants in the no standard group were highly overconfident (as in the current Experiment 1). By contrast, students' idea-unit judgments showed excellent levels of accuracy and minimal overconfidence; that is, students accurately judged which key ideas were present in their recall response and which were not (for similar results with college-age learners, see Dunlosky et al., 2011). Thus, making these idea-unit judgments may help learners to identify which material is less well known and in need of further study.

Experiment 2 included three groups. The prompts only group and the SRL support group were exactly the same as in Experiment 1. The third group was the *self-regulated learning with idea-unit judgments* group (henceforth, the *IDU group*). The procedure for this group was the same as in the SRL group, except that participants also made idea-unit judgments for each of their responses. A key prediction for Experiment 2 was that the idea-unit judgments would be highly accurate and in turn would improve learning by providing a better basis for making effective control decisions. As in Experiment 1, we conducted secondary analyses to evaluate the degree to which the idea-unit judgments are related to control decisions, to explore the intriguing possibility that this fine-grained judgments would be more highly related to control and hence to subsequent learning gains.

Method

Participants and design. Participants were 87 undergraduates from a Midwestern university who participated for course credit. The sample was 61% female, 86% Caucasian, 8% African American, 5% Latino, 1% American Indian/Alaska Native, and 3% Asian. Mean sTOFHLA score was 34.0 (SE = 0.4, range 15–36) and mean METER score was 37.2 (SE = 0.4, range 11–40), indicating that the majority of the sample had adequate health literacy.

Participants were randomly assigned to one of three groups (prompts only group, SRL support group, or IDU group).

Materials and procedure. The materials for Experiment 2 were the same as in Experiment 1, except for one alteration to the prior knowledge questionnaire. Instead of answering the eight basic questions used in Experiment 1, participants were pretested on the 12 prompts used in Sessions 2 and 3.

The procedure for Session 1 was the same as in Experiment 1. Responses to the demographic questions were as follows: *Do you have diabetes?* (of the 86 who responded, two responded yes), *If yes, how old were you when you were diagnosed?* (M = 11 years), *If yes, how old are you now?* (M = 20 years), *If yes, have you attended a patient education program in the past?* (one responded yes), and *Do you know anyone who has diabetes?* (94% responded that they had at least one friend or family member with diabetes; of those who responded yes, the average number of people known with diabetes was 2, range 1–6).

Two days after Session 1, Session 2 again involved a 50-min study period. The procedure for the prompts only group and the SRL support group were the same as in Experiment 1. The procedure for the IDU group was the same as in the SRL support group, but with one additional step. After participants made their self-score judgment for a response, the next screen displayed their response along with the key idea units in the correct answers for that particular question (for an example, see Figure 6). The directions stated: "For each answer, decide if you included it in your response above. If so, check the box next to that answer. Only check off answers that you wrote in your response on this trial (if you wrote it on some other trial but didn't write it this time, don't check it off)." Once participants checked off their answer choices, they clicked a button at the bottom of the screen to submit their idea-unit judgments. The next screen then listed the topic-relevant passages for self-paced study just as in the SRL support group (see Figure 3). This process continued until the 50-min study period was complete. Session 3 took place five days later and was the same as in Experiment 1.

Results and Discussion

Prior knowledge about diabetes. Prior knowledge responses were scored as in Experiment 1. The three groups did not differ significantly, $F < 1$ (words per question: $M = 2.8$, $SE = 0.9$, in the prompts only group; $M = 3.5$, $SE = 0.9$, in the SRL support group; and $M = 4.1$, $SE = 1.0$, in the IDU group).

Figure 6. Illustration of the screen layout used to elicit idea-unit judgments in the IDU group during Session 2 in Experiment 2. A prompt is shown at the top of the screen, along with the participant's response in an uneditable text field. The key idea units in the correct answers are shown in the list below the response, along with check boxes for participants to indicate which answers they believe they included in their response.

Final test performance. The mean number of idea units correctly recalled for each group is reported in Figure 7. A one-way ANOVA revealed a significant effect of group, $F(2, 83) = 19.28$, $MSE = 153.90$, $p < .001$, $\eta_p^2 = .32$. Follow-up t tests showed that the IDU group significantly outperformed the prompts only group and the SRL support group, $t(56) = 5.10$, $p < .001$, $d = 1.34$, and $t(55) = 6.13$, $p < .001$, $d = 1.62$, respectively. In contrast to Experiment 1, performance in the prompts only group and the SRL support group did not significantly differ, $t(55) = 0.28$, $d = 0.07$. Whereas the level of performance in the SRL support was relatively consistent across experiments, performance in the prompts only group was 6% higher in Experiment 2 than in Experiment 1. We have no explanation for this outcome. The only methodological difference between Experiments 1 and 2 concerned the nature of the prior knowledge prompts, such that the prior knowledge questions used in Experiment 2 were the same as the prompts used during restudy. Preexposure to the prompt questions during Session 1 may have alerted participants in the prompts only group to the importance of this information.

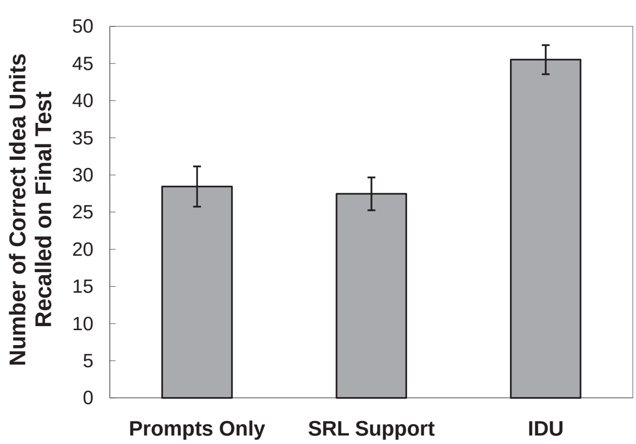

Figure 7. Mean number of idea units correctly recalled on the final test as a function of group in Experiment 2. Error bars report standard errors of the means.

Most important, the advantage of the IDU group over the other two groups was substantial (a 63% relative improvement over the other two groups). Based on the significant advantage of the IDU group, we conducted secondary analyses to evaluate the extent to which this advantage may have been attributable to better self-regulated learning.

Relative accuracy of judgments. To analyze monitoring accuracy, we again examined the relative accuracy of the self-score judgments made during Session 2. The mean across individ-

ual correlations is reported in Table 1. Although relative accuracy was significantly greater than zero for the SRL support group [$t(26) = 10.88$, $p < .001$] and the IDU group [$t(26) = 9.07$, $p < .001$], it was still only modest at best. For the IDU group, we also computed relative accuracy based on the idea-unit judgments by computing the gamma correlation between the number of idea units they said they correctly recalled and the number of idea units they actually recalled on each trial. The mean correlation across individuals (see Table 1) was greater than zero, $t(28) = 13.58$, $p < .001$. Results showed that relative accuracy was substantially greater for idea-unit judgments than for self-score judgments, $t(26) = 4.69$, $p < .001$, $d = 1.15$.

Absolute accuracy of judgments. Regarding absolute accuracy, recall responses from Session 2 were separated into the same four categories as in Experiment 1. For each individual, we then computed the mean self-score judgment for responses in each category. Means across individuals are reported in Figure 8. The SRL support group demonstrated the same qualitative trend as in Experiment 1. Self-score judgments were significantly lower for responses with 1–2 correctly recalled idea units versus 3–4 correctly recalled idea units [$t(25) = 7.66$, $p < .001$, $d = 0.84$], which in turn were lower than for responses with 5+ correctly recalled idea units [$t(20) = 2.26$, $p = .035$, $d = 0.53$]. The IDU group showed the same qualitative pattern, with lower self-score judgments for responses with 1–2 correctly recalled idea units versus 3–4 correctly recalled idea units [$t(26) = 4.74$, $p < .001$, $d = 0.76$], and for 3–4 correctly recalled idea units versus 5+ correctly recalled idea units [$t(25) = 1.18$, $p = .25$, $d = 0.28$]. Again, these results parallel the analysis of relative accuracy, with participants' judgments increasing with the amount of correctly recalled information.

Most important are analyses of judgments for commission errors. Participants were still overconfident in their commissions when making self-score judgments. In the SRL support group, self-score judgments for commissions were significantly greater than zero [$t(24) = 17.61$, $p < .001$] and did not significantly differ from self-score judgments for responses with 1–2 correct idea units [$t(24) = .48$, $p = .64$, $d = 0.07$].[2] Likewise, in the IDU group, self-score judgments for commissions were significantly greater than zero [$t(18) = 8.08$, $p < .001$] and did not significantly differ from self-score judgments for responses with 1–2 correct idea units [$t(18) = .78$, $p = .45$, $d = 0.17$]. For the IDU group, we also computed the absolute accuracy of the idea-unit judgments. Idea-unit judgments for commissions were also greater than zero [$t(20) = 2.78$, $p = .012$]. However, idea-unit judgments were significantly lower for commissions than for responses with 1–2 idea units [$t(20) = 5.16$, $p < .001$, $d = 0.59$]. Furthermore, idea-unit judgments for commissions were substantially lower than self-scores for commissions [$t(18) = 8.06$, $p < .001$, $d = 2.45$]. Thus, as evident from inspection of Figure 8, idea-unit judgments showed relatively minimal overconfidence for commission errors.

To summarize, idea-unit judgments were highly accurate, which could in turn influence control decisions. If a learner realizes that he or she does not know the answer to a specific question, the subsequent restudy decision should be to restudy the relevant information about that particular topic. Thus, we again examined the extent to which judgments were related to control decisions.

Control of restudy during Session 2. Before we examine the relationship between monitoring and control, we provide a brief summary of restudy measures in all three groups during Session 2. For each participant, we calculated how many of the 33 target passages were restudied at least once. A one-way ANOVA revealed a significant effect of group, $F(2, 84) = 12.23$, $MSE = 50.82$, $p < .001$, $\eta_p^2 = .23$. More passages were restudied at least once in the prompts only group ($M = 31.0$, $SE = 0.7$) than in the SRL support group ($M = 24.4$, $SE = 1.5$) or in the IDU group ($M = 22.1$, $SE = 1.6$), $t(56) = 4.09$, $p < .001$, $d = 1.07$, and $t(56) = 5.09$, $p < .001$, $d = 1.34$, respectively. We also calculated the total number of restudy events as in Experiment 1. Four values more than 3 SDs above the mean were removed as outliers. A one-way ANOVA revealed a significant effect of group, $F(2, 80) = 12.15$, $MSE = 597.18$, $p < .001$, $\eta_p^2 = .23$. The total number of restudy events was greater in the prompts only group ($M = 69.4$, $SE = 4.9$) than in the SRL support group ($M = 51.5$, $SE = 4.9$) or in the IDU group ($M = 36.5$, $SE = 4.2$), $t(52) = 2.58$, $p = .013$, $d = 0.70$, and $t(52) = 5.15$, $p < .001$, $d = 1.40$, respectively; the latter two groups also differed significantly, $t(56) = 2.34$, $p = .023$, 0.61. Finally, we calculated the mean amount of restudy time (in seconds) spent per restudy event for each participant. A one-way ANOVA revealed a significant effect of group, $F(2, 79) = 22.43$, $MSE = 159.29$, $p < .001$, $\eta_p^2 = .36$. Time per restudy event was greater in the prompts only group ($M = 34.2$, $SE = 3.3$) than in the SRL support group ($M = 12.3$, $SE = 1.8$) or in the IDU group ($M = 15.5$, $SE = 2.2$), $t(51) = 6.15$, $p < .001$, $d = 1.70$, and $t(51) = 4.83$, $p < .001$, $d = 1.33$, respectively; the latter two groups did not differ significantly, $t(56) = 1.12$, $d = 0.29$. Taken together, these results indicate that the advantage in final test performance for the IDU group was not attributable to participants in this group restudying more passages or restudying them longer.

Relationship between monitoring and control. To evaluate the extent to which monitoring was related to control decisions, we examined correlations between judgments and two measures of control as in Experiment 1. Because the IDU group made both self-score judgments and idea unit judgments, we report the means

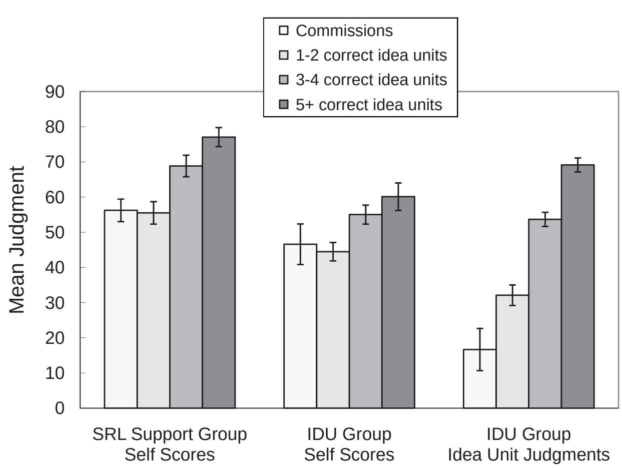

Figure 8. Mean judgments as a function of the number of correct idea units actually contained in a self-test response during Session 2 in Experiment 2. The left bars report mean self-score judgments in the SRL-support group, the middle bars report mean self-score judgments for the IDU group, and the right bars report mean idea-unit judgments for the IDU group. Error bars report standard errors of the means.

across individual correlations for both kinds of judgments in Table 2 for completeness. However, analyses for the IDU group below focus on outcomes involving the idea-unit judgments because they were temporally more proximal than self-score judgments to control decisions in this group.

First, in the SRL-support group, the correlation between self-score judgments and number of passages studied during restudy opportunities was less than zero, $t(26) = 3.02$, $p = .006$. In the IDU group, the correlation between idea-unit judgments and number of passages studied during restudy opportunities was less than zero, $t(27) = 6.66$, $p < .001$. This correlation was stronger in the IDU group than in the SRL support group, $t(53) = 3.82$, $p < .001$, $d = 1.03$.

Second, in the SRL support group, the correlation between self-score judgments and total time spent studying the selected passages was less than zero, $t(26) = 2.59$, $p = .016$. In the IDU group, the correlation between idea-unit judgments and total time spent studying the selected passages was also less than zero, $t(27) = 6.71$, $p < .001$. Once again, the correlation was stronger in the IDU group than in the SRL support group, $t(53) = 3.92$, $p < .001$, $d = 1.06$.

In sum, as learners' judgments decreased, the number of total passages studied increased and the total time studying passages increased. As in Experiment 1, results of these secondary analyses suggest that the restudy decisions of participants were at least partially guided by their metacognitive judgments. Most important, not only was accuracy substantially greater for idea-unit judgments than for self-score judgments (in terms of both relative and absolute accuracy), the idea-unit judgments were also more closely linked to control decisions.

Relationship between control decisions and learning gains. Finally, we examined the relationship between control decisions during study and learning gain, as in Experiment 1. We computed gamma correlations for learning gain and total number of passages studied and for learning gain and total time studying all passages for each individual. Means across individual correlations are reported in Table 3. One-sample *t* tests showed that means were not greater than zero for the SRL support group [$t(25) = .36$, $p = .723$, and $t(25) = .13$, $p = .898$, respectively]. One possibility for why these correlations were weaker than in Experiment 1 may have to do with the related finding that the associations between control decisions and monitoring were also weaker for the SRL group in Experiment 2 than in Experiment 1 (see Table 2), suggesting that participants in Experiment 2 were less effective at self-regulating their restudy. If so, a weaker association between control and learning gain would be expected. More important, both correlations were significantly greater than zero for the IDU group [$t(27) = 4.15$, $p < .001$, and $t(27) = 4.78$, $p < .001$, respectively]. These results confirm that control decisions during restudy were related to learning gains for learners in the IDU group.

General Discussion

Effective management of chronic diseases such as diabetes can depend on the extent to which patients can learn and remember disease-relevant information. Whereas most prior research directed at improving people's learning of patient education materials has focused on revising the content or display of the materials, the current study is the first to explore a complementary approach that targets learners' self-regulatory processes. Specifically, we explored the promise of an educational technology based on the dual power of self-testing (Rawson & Dunlosky, in press): Self-testing has an indirect effect on learning through improving monitoring and control processes and can also exert a direct influence on learning via boosting memory when retrieval is successful (Roediger & Karpicke, 2006). Although results from Experiment 1 were promising, the effects of providing support for self-regulated learning (via self-testing) had a modest effect. One difficulty may have been that self-testing alone did not yield very accurate monitoring. Self-score judgments showed low levels of relative accuracy (0.33) and substantial overconfidence in commission errors. Accordingly, in Experiment 2, self-testing was paired with a standard of evaluation that yielded highly accurate monitoring (i.e., idea-unit judgments), and this enhanced support for monitoring produced a substantial increase in performance (63%) relative to baseline performance.

As important, results from both experiments provide insight into why this particular technique benefited learning, because both the SRL support group and the IDU group were prompted to test themselves on the target material. If the present benefit of self-testing was attributable entirely to its *direct* influence on performance, these two groups should have performed similarly and better than the prompts-only group. However, the IDU group significantly outperformed the SRL support group, which did not consistently outperform the prompts only group. Thus, although self-testing can have dual benefits for learning and retention, direct effects of the act of retrieval itself do not appear to be the primary reason why the current technique improved people's learning of patient education materials (although we cannot rule out the possibility that the IDU judgments may have influenced retrieval by encouraging more effortful or elaborative retrieval attempts, which can facilitate learning; see Carpenter, 2009; Pyc & Rawson, 2009).

Another reason why the IDU group outperformed the other groups might have to do with the nature of the feedback inherently provided by the IDU judgment prompts, in that they provide a summary list of the correct answers. For instance, the feedback provided by the IDU judgment prompt may have increased the likelihood that participants set more specific goals for learning, and setting specific goals has been shown to improve performance (for a review, see Locke & Latham, 2006). Although further research would be needed to conclusively determine the relative contributions of IDU feedback versus participants' self-regulated restudy to the learning advantage in the IDU group, some outcomes here suggest that the effect is not entirely attributable to IDU feedback. First, participants in the IDU group engaged in a considerable amount of restudying after making their judgments, and it seems unlikely that this self-regulated restudy contributed nothing above and beyond the more minimal feedback provided in the IDU prompt. Second, the significant correlations between learning gain and control in the IDU group (see Table 3) suggest that learning was related to self-regulated study. Thus, the advantage in the IDU group appears to be attributable at least in part to the enhanced monitoring and control processes afforded by idea-unit judgments.

Idea-unit judgments were markedly more accurate than the corresponding self-score judgments, in terms of both relative and absolute accuracy, and their high level of accuracy does appear to improve people's learning of patient education materials. This benefit may partly arise because the judgments provide accurate input that partic-

ipants can use to decide which materials to study. Indeed, the strongest association between monitoring judgments and decisions about what to restudy were observed in the IDU group in Experiment 2: When participants indicated that they had recalled fewer idea units from a passage, they were more likely to restudy the passages (−.38) and spent more time studying them (−.35).

Although these results indicate that learners tended to focus on passages they judged that they could not recall well, did this strategy actually benefit performance? To provide a preliminary answer to this question, consider the correlation between (a) learning gain and (b) the strength of the association between monitoring and control across individuals in the SRL support and IDU groups in Experiment 2. Individuals who primarily choose to restudy material that they judge as less well learned show strong negative correlations between judgments and control (number of passages selected, total restudy time). If this strategy is effective, these individuals should also show greater learning gains (cf. analyses reported by Thiede, 1999). Collapsing across the SRL support group and the IDU group, the correlation between individuals' monitoring-control association and learning gain was $r = -0.44$ when the control measure was number of passages selected and $r = -0.44$ for total study time. Thus, individuals with a stronger negative monitoring-control association showed greater learning gains. Even when we examine these correlations just across participants in the idea-unit group, the correlations are still significant despite the smaller sample size, $r = -0.32$ and $r = -0.36$.

The present results are consistent with the core assumptions of general theories of self-regulated learning: Effective monitoring supports effective control, which in turn enhances learning. Despite the intuitive appeal of these assumptions, few studies have simultaneously empirically evaluated these relationships. Typically, only one of these relationships is investigated in any one study, although Thiede (1999) did report both of the expected relationships when students were learning Swahili-English word pairs (see also, Kornell & Metcalfe, 2006). Accordingly, the current study provides a critical demonstration that core assumptions of self-regulated learning theories generalize to people's learning of lengthy, difficult, and representative text materials (nearly all studies examining judgments and control of text learning have participants study multiple short texts on unrelated topics, which is not representative of the kinds of materials people usually study).

The cascading effect of monitoring accuracy on subsequent control processes and learning has important implications for improving patient education materials. It is evident that without the support of idea-unit standards, self-testing alone can lead to relatively inaccurate judgments of conceptual materials (see also, Dunlosky et al., 2005). Given that pairing self-testing and idea-unit judgments yielded better accuracy and improved learning in the present experiments, we encourage further exploration of this new kind of monitoring judgment and cautiously advocate its use in improving patient education materials. Importantly, this recommendation is not intended to supplant efforts to improve materials by enhancing readability or other text characteristics as has commonly been done in prior research. Rather, we see these as complementary approaches that together may lead to the most successful learning outcomes. Indeed, an intriguing possibility is that these two factors may interact, such that supports for self-regulated learning may be less effective when incorporated into poorly designed materials.

One factor for further exploration concerns the mode of presentation. In the present context, we used a computer interface to present the materials and to prompt self-testing and idea-unit judgments. These techniques could be easily transferred to written materials. For instance, each of the NIDDK booklets from which the current materials were drawn could include a supplemental page that included prompt questions for self-testing over key content. A separate page could provide an "answer key" with the correct idea units for checking responses, along with page numbers indicating where in the booklet the relevant information is contained to facilitate restudy. Another important direction for further exploration concerns generalizing these promising findings to other populations. A potential limitation of the current study is that these experiments involved healthy undergraduates rather than patients with diabetes. Whereas our student sample tended to have minimal prior knowledge about diabetes, patients would likely have more variable levels of prior knowledge about the target material, which may influence metacognitive processes (e.g., by producing overconfidence in current knowledge and thus reduced effort to study material). Consistent with this possibility, Brown and Park (2002) reported that younger and older adults learned less new information from a text about a familiar disease than from a text about an unfamiliar disease. Furthermore, individuals with diabetes on average would be expected to be older and have a lower level of literacy than college students, which raises the possibility that they may monitor less accurately even with support via idea-unit judgments. Of indirect relevance, Lipko et al. (2009) showed that middle-school students can accurately evaluate their own recall responses when using idea-unit judgments, suggesting that this form of support for self-regulated learning may be effective for populations with lower levels of functioning than college students. Finally, another key difference between college students and diabetes patients may involve motivation. An intuitive assumption is that patients who have been diagnosed with diabetes would be more motivated than students to learn information about the disease. If so, the current results may be particularly promising by showing that support for self-regulated learning can enhance learning even for less motivated readers.

In conclusion, the current study provides an important first step toward demonstrating the promise of incorporating support for self-regulated learning into patient education materials, and further exploration of the benefits for patient learning will be a fruitful direction for future research.

References

Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., & Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Education and Counseling, 38,* 33–42.

Baker, G. C., Newton, D. E., & Bergstresser, P. R. (1988). Increased readability improves the comprehension of written information for patients with skin disease. *Journal of the American Academy of Dermatology, 19,* 1135–1141.

Barat, I., Andreasen, F., & Damsgaard, E. M. S. (2001). Drug therapy in the elderly: What doctors believe and patients actually do. *British Journal of Clinical Pharmacology, 51,* 615–622.

Brown, S. C., & Park, D. C. (2002). Roles of age and familiarity in learning health information. *Educational Gerontology, 28,* 695–710.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569.

Cegala, D. J. (1997). A study of doctors' and patients' patterns of information exchange and relational communication during a primary care consultation: Implications for communication skills training. *Journal of Health Communication, 2,* 169–194.

Cotugna, N., Vickery, C. E., & Carpenter-Haefele, K. M. (2005). Evaluation of literacy level of patient education pp. in health-related journals. *Journal of Community Health: The Publication for Health Promotion and Disease Prevention, 30,* 213–219.

Davis, T. C., Crouch, M. A., Wills, G., Miller, S., & Abdehou, D. M. (1990). The gap between patient reading comprehension and the readability of patient education materials. *Journal of Family Practice, 31,* 533–539.

Dowe, M. C., Lawrence, P. A., Carlson, J., & Keyserling, T. C. (1997). Patients' use of health-teaching materials at three readability levels. *Applied Nursing Research, 10,* 86–93.

Dunlosky, J., Hartwig, M., Rawson, K. A., & Lipko, A. R. (2011). Improving college students' evaluation of text learning using idea-unit standards. *Quarterly Journal of Experimental Psychology, 64, 467*–484.

Dunlosky, J., Hertzog, C., Kennedy, M. R. T., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10,* 4–11.

Dunlosky, J., Rawson, K. A., & McDonald, S. L. (2002). Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T. Perfect & B. Schwartz (Eds.), *Applied metacognition* (pp. 68–92). Cambridge, UK: Cambridge University Press.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. To appear in B. Ross (Ed.), *Psychology of Learning and Motivation, 54,* 103–140.

Estey, A., Musseau, A., & Keehn, L. (1994). Patients' understanding of health information: A multihospital comparison. *Patient Education & Counseling, 24,* 73–78.

Gal, I., & Prigat, A. (2005). Why organizations continue to create patient information leaflets with readability and usability problems: An exploratory study. *Health Education Research, 20*(4), 485–493.

Gazmararian, J. A., Williams, M. V., Peel, J., & Baker, D. W. (2003). Health literacy and knowledge of chronic disease. *Patient Education and Counseling, 51,* 267–275.

Gordon, H. S., Street, R. L., Jr., Kelly, P. A., Souchek, J., & Wray, N. P. (2005). Physician-patient communication following invasive procedures: An analysis of post-angiogram consultations. *Social Science & Medicine, 61,* 1015–1025.

Hoffmann, T., & McKenna, K. (2006). Analysis of stroke patients' and carers' reading ability and the content and design of written materials: Recommendations for improving written stroke information. *Patient Education and Counseling, 60,* 286–293.

Jacobson, T. A., Thomas, D. M., Morton, F. J., Offutt, G., Shevlin, J., & Ray, S. (1999). Use of a low-literacy patient education tool to enhance pneumococcal vaccination rates: A randomized controlled trial. *Journal of the American Medical Association, 282,* 646–650.

Kenny, T., Wilson, R. G., Purves, I. N., Clark, J., Newton, L. D., Newton, D. P., & Moseley, D. V. (1998). A PIL for every ill? Patient information leaflets (PILs): A review of past, present and future use. *Family Practice, 15,* 471–479.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 609–622.

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied, 15,* 307–318.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15,* 265–268.

Ong, L. M. L., de Haes, J. C. J. M., Hoos, A. M., & Lammes, F. B. (1995). Doctor-patient communication: A review of the literature. *Social Science & Medicine, 40,* 903–918.

Perkins, L., & Cohen, J. (2008). Meeting patient needs in the hospital setting – are written nutrition education resources too hard to understand? *Nutrition & Dietetics, 65,* 216–221.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385–407.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447.

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology 19,* 559–579.

Rawson, K. A., & Dunlosky, J. (in press). Retrieval-Monitoring-Feedback (RMF) technique for producing efficient and durable student learning. To appear in R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies.* Springer.

Rawson, K. A., Gunstad, J., Hughes, J., Spitznagel, M. B., Potter, V., Waechter, D., & Rosneck, J. (2009). The METER: A brief, self-administered measure of health literacy. *Journal of General Internal Medicine, 25,* 67–71.

Rhodes, M., & Tauber, S. (in press). The influence of delayed judgments of learning (JOLs) on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin.*

Roediger, H. L., III, & Butler, Andrew, C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27.

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210.

Sheard, C., & Garrud, P. (2006). Evaluation of generic patient information: Effects on health outcomes, knowledge and satisfaction. *Patient Education and Counseling, 61,* 43–47.

Singh, J. (2000). The readability of HIV/AIDS education materials. *AIDS Education and Prevention, 12,* 214–224.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 204–221.

Suhonen, R., & Leino-Kilpi, H. (2006). Adult surgical patients and the information provided to them by nurses: A literature review. *Patient Education and Counseling, 61,* 5–15.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662–667.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73.

Vallance, J. K., Taylor, L. M., & Lavallee, C. (2008). Suitability and readability assessment of educational print resources related to physical activity: Implications and recommendations for practice. *Patient Education and Counseling, 72*(2), 342–349.

White, S., & Dillow, S. (2005). *Key concepts and features of the 2003 National Assessment of Adult Literacy* (NCES 2006–471). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Williams, M. V., Baker, D. W., Honig, E. G., Lee, T. M., & Nowlan, A. (1998). Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest, 114,* 1008–1015.

Wilson, E. A. H., & Wolf, M. S. (2009). Working memory and the design of health materials: A cognitive factors perspective. *Patient Education and Counseling, 74*(3), 318–322.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ, US: Erlbaum Publishers.

Winslow, E. H. (1998). Caring for patients with limited literacy. *American Journal of Nursing, 98,* 55–57.

Wolf, M. S., Davis, T. C., Shrank, W. H., Neuberger, M., & Parker, R. M. (2006). A critical review of FDA-approved Medication Guides. *Patient Education and Counseling, 62,* 316–322.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25,* 3–17.

Received December 30, 2010
Revision received May 25, 2011
Accepted June 1, 2011 ■