Running Head: EFFECTIVE METACOGNITION

Effective Implementation of Metacognition

Michael J. Serra and Janet Metcalfe

Columbia University

Effective Implementation of Metacognition

Metacognition can be described as "a critical analysis of thought," "knowledge and cognition about cognitive phenomena" (Flavell, 1979), or simply "thinking about thinking." It can focus on any aspect of cognition, even metacognition itself (Dunlosky, Serra, Matvey, & Rawson, 2005; Nelson & Narens, 1994). Metacognition has typically been conceptualized as involving one or more of the following aspects of a cognitive process: knowledge about that process, the monitoring of that process, and the control of that process. When optimized, these aspects of metacognition can augment performance of the target cognition, including students' learning (Azevedo & Cromley, 2004; Winne, 1995). For this reason, many researchers have sought to train students to engage in metacognitive thinking to improve their learning (e.g., White & Frederiksen, 1998).

Unfortunately, metacognition is prone to errors and often exhibits poor accuracy, which can have a negative impact on its implementation in studying and learning (e.g., Nelson & Dunlosky, 1991; Maki, 1998b; Metcalfe & Finn, 2008; Thiede, 1999; Thiede, Anderson, & Therriault, 2003). Although we believe that metacognition should be taught in the classroom as an important learning skill, students and teachers must be made aware of the errors that can accompany metacognition and learn ways to sidestep them. Towards this end, this chapter discusses the effective implementation of metacognition. We present an overview of metacognition and discuss how its three main components—knowledge, monitoring, and control—play a role in studying for retention and comprehension. For each component, we review some of the literature relevant to optimizing metacognition. In particular, we focus on errors/biases in monitoring, and the negative effects that poor monitoring accuracy can have on control processes. We also review some of the heuristics that people use to make their metacognitive judgments, and methods that have been shown to improve the accuracy of metacognitive judgments. We also suggest some ways to best incorporate accurate monitoring into study.

Knowledge

Metacognitive knowledge is information that one consults when thinking about a particular cognition. This knowledge can include information about the cognitive task at hand, about one's ability to perform that task, or about potential strategies one might use to perform that task. Metacognitive knowledge informs monitoring and control, and the accuracy and efficacy of these latter two functions increase as metacognitive knowledge increases (Winne & Hadwin, 1998, see also Wilson & Brekke, 1994). Metacognitive knowledge might exert an influence in the classroom in a number of ways. For example, a student who expects an essay test instead of a multiple-choice test on a chapter might study by writing a summary of the chapter instead of attempting to answer the practice short-answer questions in the back of the chapter. He or she presumably knows that trying to write a summary is more representative of how understanding will be tested on such a test than answering the multiple-choice questions is (knowledge affecting monitoring). Another student who knows that he or she struggles to understand mathematics lessons might plan to devote extra time to study or seek extra help (knowledge affecting control).

Students who know more about how to study and about how learning occurs (i.e., those with more metacognitive knowledge) learn better than those with less metacognitive knowledge (Winne & Hadwin, 1998). For this reason, educating students about how they learn and identifying effective (and ineffective) learning strategies for them should not only improve the accuracy of their metacognitive judgments, but should also improve their self-regulated learning (Azevedo & Cromley, 2004; White & Frederiksen, 1998). Teachers and researchers can refer to the practice guides available on the Institute of Education Sciences (IES) website for helpful reviews of the efficacy of approaches to various educational topics (e.g., encouraging girls in math and science) to become better informed themselves in terms of which instructional methods are effective and which are not. Of most relevance to the current topic will be the guide on organizing instruction and study to improve student learning (Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel, & Metcalfe, 2007). This guide discusses the efficacy of various instructional approaches and techniques that teachers and students might not be aware of, such as the benefits of spacing learning, effective study time allocation, and the benefits of testing for subsequent retention.

In terms of improving students' metacognition, teachers should warn students about metacognitive illusions (i.e., systematic errors in metacognitive monitoring) they might encounter when studying specific types of materials or making specific types of judgments (e.g., the foresight bias discussed later). Although students can learn to overcome metacognitive illusions through experience with the illusion-causing materials, only explicit, theory-based training for such illusions transfers to new instances of such materials (Koriat & Bjork, 2006). For this reason, students should be explicitly trained to avoid specific cognitive illusions (e.g., illusions of knowing, use of familiarity heuristics) rather than waiting for them to learn from their mistakes. Further empirical work is certainly required here as simply teaching or warning students about metacognitive illusions is not adequate to ensure that they will not occur. As described by Wilson and Brekke (1994), such "mental contamination" can be difficult to avoid if one does not feel that it is in fact occurring. Not only will students need to accept that they are prone to specific biases, but they must also learn how to appropriately adjust their metacognitive monitoring judgments to avoid them. Sometimes, however, the metacognitive question can be framed in such a way that the student will recognize the possibility that they might experience a metacognitive illusion. For example, students who are asked whether they will *remember* information on a test will have a different response than students who are asked whether they will *forget* information on a test (Finn, 2008). Framing the question in a different way can facilitate the insight that the student's metacognition might be incorrect.

Monitoring

Metacognitive monitoring focuses on the progress of the cognitive process in which the person is engaged. Such monitoring can take the form of explicit judgments as are typically elicited in laboratory studies of metacognition. These judgments can be made in a simple YES/NO fashion, but are often requested on a continuous scale (e.g., 1-7, 0-100). Numerous monitoring judgments have been examined (e.g., feeling-of-knowing judgments or FOKs, ease-of-learning judgments or EOLs, confidence-in-response judgments, etc.), but this chapter will mainly focus on two: (1) judgments of learning (JOLs), which evaluate one's memory, and (2) metacomprehension judgments, which evaluate one's understanding of text materials. Although some judgments were initially conceptualized as having direct access to cognitive processes (e.g., Hart, 1967) that position was abandoned in light of evidence suggesting that monitoring judgments are inferential in nature (e.g., Vesonder & Voss, 1985). Rather than having direct access to the cognitive process being evaluated, metacognitive monitoring judgments are made inferentially based on cues that are related to that process (Koriat, 1997). For instance, students studying for a test must make an inference about their learning of the information based on aspects of the materials being studied (e.g., if a chapter in a textbook was written unclearly) or about their experience in studying the materials (e.g., if the information seemed easy to understand). These inferences are also informed by metacognitive knowledge about oneself or the task being judged (Koriat, 1997), such as the knowledge that essay tests require in-depth studying or the belief that one is knowledgeable about physics.

Much research on metacognition has focused on the accuracy of monitoring judgments, which has mainly been conceptualized in two ways: calibration and relative accuracy. The *calibration* of one's judgments refers to a difference score between the mean of one's predictive judgments and one's performance on the task being judged or predicted. Consider hypothetical participants in a laboratory study on metamemory (metacognition about memory). While studying paired associate items (e.g., two unrelated nouns) for the criterion test, the participants make JOLs on a 0%-100% scale indicating the percent-likelihood that they will correctly remember each item on a test. Suppose that the overall mean of their JOLs is 88%. Assuming that the participants correctly remember a mean of 66% of the items on the test, their calibration score will be +22%, indicating overconfidence. A group of participants in such a study would be said to demonstrate "good" calibration if the overall mean of their JOLs did not significantly differ from the overall mean of their performance scores.

The *relative accuracy* of one's judgments refers to a measure of how well one's judgments differentiate performance on the cognitive task being judged. This measure is usually calculated by computing a gamma correlation between one's judgments about individual items and performance on those same test items. Like a Pearson correlation, gamma can range from –1.0 to +1.0. If a hypothetical participant gives mostly high JOLs to items they will remember on the test and low JOLs to items they will not remember on the test, their gamma correlation will be positive. Doing the opposite will result in a negative correlation. Assigning JOLs to items at random tends to result in a correlation of zero (or a correlation might even be incalculable). Typically, a gamma correlation is calculated for each participant. The mean gamma correlation is then calculated across all of the participants in a group to estimate the relative accuracy of their judgments; if the mean is significantly greater than zero, the group's judgments are said to be above chance.

Metacomprehension research tends to focus on the relative accuracy of metacomprehension judgments rather than on their calibration. In fact, the relative accuracy of metacomprehension judgments is often referred to simply as *metacomprehension accuracy*. Although metacomprehension judgments typically demonstrate above-chance relative accuracy, it is typically low (i.e. gamma correlations around or less than .3, see Maki, 1998b; Weaver, Bryant, & Burns, 1995). Because of the importance of these self-assessments and because the correlations are typically unacceptably low, much research on metacomprehension judgments has focused on ways to make them more accurate.

It should be noted that the two measures of accuracy discussed above are not always in agreement and do not represent the accuracy of the judgments in the same way (Nelson, 1996); judgments might be considered accurate for one of the measures and inaccurate for the other. For example, Koriat, Sheffer, & Ma'ayan (2002) demonstrated the underconfidence-with-practice (UWP) effect in which learners' JOLs become underconfident from a second study-test trial on, indicating poor calibration. The underconfident judgments in Koriat et al.'s (2002) study, however, demonstrated good *relative* accuracy that actually increased across trials.

The accuracy of metacognitive judgments would not matter if metacognition were epiphenomenal. As we discuss later, however, monitoring plays a role in the control of the cognitive process being evaluated (Finn, 2008; Metcalfe & Finn, 2008; Thiede et al., 2003).

Because of this connection, the efficacy of that control will often be linked to the accuracy of the monitoring judgments (Thiede, 1999; Thiede et al., 2003).

Heuristics that Can Produce Inaccuracies in Metacognitive Monitoring

Although both JOLs and metacomprehension judgments typically demonstrate above-chance relative accuracy, the calibration and relative accuracy of these judgments is not always impressive (Maki 1998b, Nelson & Dunlosky, 1991; Weaver & Kelemen, 1997). In many situations, the low predictability of these judgments can be attributed to participants' basing them on information that is not diagnostic of their future performance. As described by Metcalfe, Schwartz, and Joaquin (1993):

> [When] making judgments related to the external world rather than to their internal states and abilities, people use other heuristics or rules of thumb. These heuristics approximate the uncertain quantity indirectly, rather than measuring the quantity itself. Because they do not measure the quantity directly, such heuristics may result in biases and errors. (p. 860).

In other situations (e.g., delayed JOLs, as we discuss later), calibration and relative accuracy can be quite good because participants base their judgments on information that *is* diagnostic of their future performance. This section reviews some heuristics that are sometimes used to inform metacognitive judgments and that can produce errors and illusions in monitoring. This section is not meant to describe *all* of the potential errors and poor heuristics that might arise in metacognitive monitoring, but simply to provide some illustrative examples. It also suggests some ways that the errors associated with these particular cues might materialize in the classroom. As we learn more about heuristics that lead to accurate and inaccurate metacognitive judgments, we can better understand which heuristics aid or hinder metacognitive accuracy.

Familiarity heuristics. Learners' familiarity with the information being judged can have an influence on their metacognitive judgments. Because this experience of familiarity often results from prior exposures to and learning of the information, it can be diagnostic of a greater likelihood that the information is known (i.e., it often results in accurate judgments). Unfortunately, familiarity can also arise in situations where it is unrepresentative of knowledge. For example, participants in a study by Reder and Ritter (1992) solved difficult arithmetic problems. After each was presented, they had to quickly choose whether to calculate the answer to each problem or recall it from memory (calculation, of course, was the only option the first time a problem was presented). Participants received 50 points for correctly recalling an answer and 5 points for correctly calculating an answer (but only if the selections and responses were made within the designated time limits). Participants were later paid .05 cents for each point earned. By manipulating the occurrence of specific numbers in the set of problems, Reder and Ritter manipulated the participants' familiarity for the numbers present in the problems independent of their memory for specific problems and their solutions. Participants were able to use their familiarity with the numbers—both independently and as whole problems—to quickly decide whether they knew the answers to the problems. This strategy proved helpful when the specific problem had actually been presented in accordance with the participants' familiarity for the numbers, but proved to be faulty when familiar numbers were combined into novel problems—problems for which the participants could not actually recall an answer.

Although familiarity might sometimes be an appropriate basis for metacognitive judgments, in many situations inappropriate use of familiarity during study can produce monitoring errors. For example, many textbooks provide a list of key terms that students should be knowledgeable about and practice questions that they should be able to answer after reading a chapter. A student who bases a judgment of their memory or understanding for the chapter on their familiarity with these terms or the phrases in the questions rather than on an attempt to define the terms or answer the questions would be likely to overestimate their learning and might be expected to have poor relative accuracy. Such a situation was illustrated well by Metcalfe, Schwartz, and Joaquin (1993), who manipulated the familiarity of cues for feeling of knowing judgments (aka FOKs, which are judgments that one will be able to recognize the correct target word of an incorrectly-recalled cue-target pair at a later time) independent of the retrievability of the targets. This was achieved by manipulating the frequency of the cue words across pairs using a classic interference design (i.e., A-B A-B; A-B' A-B; A-D A-B; and C-D A-B pairs). So, for the A-B A-B procedure, participants studied the same cue-target pairs on each of two lists, whereas for the A-D A-B procedure, participants studied the same cues on each of two lists, but they were matched with different targets on each list (hence a change from D to B). Participants gave higher FOKs to items that had familiar cues (e.g., A-B A-B; A-B' A-B; and A-D A-B pairs) than to pairs that had less familiar cues (C-D A-B) even though interference was produced in some of the former conditions, which impaired recognition performance.

Familiarity can also cause monitoring problems when learning must be monitored across a variety of topics, as is common in educational settings (e.g., studying for multiple final examinations during the same week). Students vary in their knowledge and understanding of the different topics they encounter in school and this information—their *domain familiarity*—is sometimes used to inform students' metacognitive judgments. Participants in a study by Glenberg, Sanocki, Epstein, and Morris (1987) read and judged their understanding for texts on a number of topics before being tested on their understanding of the texts. Glenberg et al. (l987) found that participants used their familiarity with the topic of each text as the basis for their metacomprehension judgments rather than basing their judgments on their understanding of each particular text. Basing judgments on domain familiarity rather than an understanding of the texts impaired the accuracy of the participants' judgments. Students study topics in multiple domains at the same time in the classroom, and so they might fall prey to a similar error. For example, a student studying for both a mathematics and social studies test the following week might devote more time to studying mathematics because he or she feels less knowledgeable about this topic than social studies. This might often be a good strategy, but if in fact the student has poor understanding for the current social studies lesson and does not factor this into their plan of study, they might go into the social studies test under-prepared.

Fluency heuristics. Several studies have demonstrated that judgments of learning are sometimes made based on retrieval fluency—the speed and probability with which information is brought to mind (Benjamin, Bjork, & Schwartz, 1998). Retrieval fluency is predictive of memory performance in many situations (e.g., Serra & Dunlosky, 2005), so using it to inform metacognitive judgments will often be appropriate. It has recently been suggested that the utility of cues such as retrieval fluency and ease of learning are automatically incorporated into metacognitive judgments based on their ecological validity at the time the judgment is being made (Koriat, 2008). For example, Koriat (2008) recently described the easily learned, easily remembered (ELER) heuristic, which stems from the observation that items that are easily learned are typically remembered better than items that are difficult to learn. In the studies reported by Koriat (2008), participants apparently used ease of learning as a cue to inform their metacognitive judgments in a way that was directly correlated with the validity of that cue.

The use of the fluency heuristic as a basis for metamemory judgments in situations in which it is not diagnostic of performance, however, can produce illusions in metacognitive monitoring. For example, Benjamin et al. (1998) purposely chose a task for which retrieval fluency would not be predictive of later test performance. Specifically, participants answered trivia questions (e.g., Who was the second president of the United States?) and then were asked if they could freely recall their answer to the questions, in the absence of the original questions, sometime later. Although participants judged the easiest-recalled answers as being the ones they would be the most likely to free-recall later, the most-difficult to recall answers were in fact most likely to be free-recalled, presumably because more effortful retrieval attempts were more memorable than less effortful attempts. A crucial part of the free recall task was recalling the cues themselves; the cues on which the person spent the most time (and hence were most strongly encoded) were those with difficult to retrieve targets. As a result, participants' reliance on retrieval fluency as a basis for their judgments resulted in the judgments being negatively correlated with criterion test performance.

There are also circumstances in which metacognitive judgment accuracy can be impaired by using the fluency of item processing as the basis for the judgments. Interestingly, though, warning people about the possibility of bias can help to offset it. In a study by Jacoby and Whitehouse (1989), participants first studied a list of words. New words were then added to the list and all of the words were presented to the participants, who had to identify each as being either "old" or "new". Prior to this recognition test, some of the new items were primed. Half of the participants were warned about the prime and half were not. Although priming new items increased the processing fluency of the primed items compared to the unprimed items (regardless of the participants' warning condition), the warning did have an effect on whether the items were perceived as new or old. Participants who did not receive the warning were more likely than the warned participants to judge the new, primed items to be old items. Participants who were warned about the prime were more likely to discount the increased fluency of the new, primed items and to correctly identify them as new items. This study demonstrates that some metacognitive errors can be avoided if the potential error is known about beforehand.

Current-knowledge heuristic. Sometimes, after information has been obtained or understood, people think they knew or understood it all along. Participants in a study by Fischhoff (1975) read passages detailing outcomes associated with historical events (such as a battle). They then judged if they would have predicted the outcome before reading the passage. Fischhoff demonstrated that these participants could not avoid using their knowledge of the outcome when making this judgment; they even judged that they would have correctly predicted highly unlikely events (some of which were false). The participants in his study demonstrated *hindsight bias*—a tendency to use new knowledge when thinking about the past—without even knowing that they were doing so. More so than students, *teachers* should be wary of a form of hindsight bias when judging if their students understand the lesson being taught. Teachers might overestimate their students' understanding of a lesson because—to the teacher—the lesson is easy to understand. Teachers should favor objective measures of their students' understanding such as quizzes instead of relying on their own or their students' subjective assessments of the students' learning.

Association heuristic. Some information seems easier to understand or remember when it is studied than it will actually be to remember or apply later on a test. Koriat and Bjork (2006) termed such an illusion *foresight bias* and demonstrated a type of paired-associate that produces such an effect. These pairs were composed of two words that had a strong backwards association but a weak forward association (i.e., one would be likely to think of the first word when shown the second word but not likely to think of the second word when shown the first word). For example, consider the pair “fire—blaze”. The word “blaze” is almost always freely-associated to the word “fire”, but “fire” rarely—if ever—is freely-associated to the word “blaze”. When such pairs are studied and judged in a typical metamemory procedure, the presence of both words at study makes them seem highly related. At test, however, the stimulus word (fire) is not actually likely to produce the response word (blaze). The association strength present at study produces the illusion that the response word will easily be recalled at test. Such an illusion might occur with more naturalistic materials such as when students learn foreign language vocabulary. An example would be how the apparent similarity between some English words and their Spanish translations (e.g., “computer” and “computadora”) might cause students to experience a foresight bias when studying the words. Although the two words appear similar, this similarity might not be predictive of memory for the specific Spanish translation when only the English word is shown.

Heuristics that can cause illusions of knowing. One’s experience with learning materials sometimes causes the illusion that the materials have been understood when in fact they have not. Participants in a study by Glenberg, Wilkinson, and Epstein (1982) demonstrated an illusion of knowing (i.e., their judgments were overconfident) when asked to rate their comprehension for texts containing factual contradictions. Participants often failed to find these contradictions yet rated their understanding of the texts as being high. This even occurred when factual contradictions were in two adjacent sentences (Glenberg et al., 1982). These findings suggest that readers do not attempt to monitor their understanding across a whole text, but rather at lower levels such as at the per-sentence level. Although educational materials would not be as likely as Glenberg et al.’s (1982) materials to contain factual contradictions, the findings nevertheless suggest that readers might also fail to detect contradictions in their own interpretations of text materials. They might experience that they have understood each sentence well independently, but fail to detect that they have failed to understand the text as a whole, or that their own interpretations of two related sentences are in conflict. As we will discuss later, methods that enhance metacomprehension accuracy tend to focus the reader on their overall understanding of the materials (e.g., Thiede & Anderson, 2003), which is what seems to have been lacking in this example.

Participants in a study by Dunlosky, Rawson, and Middleton (2005) also made metacomprehension judgments that demonstrated an illusion of knowing. After reading the texts they would be tested on, the participants attempted to answer several questions testing specific pieces of information from each text. They rated their understanding for each piece of information after attempting to answer each question. The participants then completed a criterion test directed at the same information. Judgments made for correct pre-judgment responses tended to be underconfident. More importantly, judgments made for pre-judgment commission errors and partially correct responses demonstrated an illusion of knowing: they were overconfident. The authors interpreted their findings as suggesting that such metacomprehension judgments reflect the quantity and quality of information recalled during the pre-judgment recall attempt. Without feedback that some of their answers were incorrect, the students judged any answer they gave (even incorrect ones) to be indicative of their understanding for the critical information (Rawson & Dunlosky, 2007).

A form of illusion of knowing can also affect JOLs, particularly when they are made immediately after study (Nelson & Dunlosky, 1991). Consider a participant who is asked to study several paired associates and judge the likelihood that they will remember each item on a test fifteen minutes later. If the judgment is to be made immediately after each item is studied, participants will be likely to experience an illusion of knowing because the item will still be in working memory and it will be difficult to evaluate the memorability of the item if it is so strongly represented (Nelson & Dunlosky, 1991; 1992; Nelson, Narens, & Dunlosky, 2004). We will discuss this issue further in the next section when we discuss a way to circumvent this specific problem (i.e., by delaying JOLs).

Heuristics that Can Improve the Accuracy of Metacognitive Monitoring

Given the importance of accurate monitoring in controlling study and the numerous factors that might cause monitoring to be less than optimal, it should not be surprising that much research in the past decade has focused on ways to improve the accuracy of metacognitive judgments. In this section we review some factors and methods that improve the accuracy of JOLs and metacomprehension judgments and discuss research that has sought to determine why accuracy is improved under these conditions.

Debiasing incorrect heuristics. As described in the previous section of this chapter, metacognitive judgments are prone to errors, biases, and metacognitive illusions. Experience with and information about these illusions can help to reduce some of these biases. Koriat and Bjork (2006) described one such illusion—foresight bias—in which some to-be-studied pairs of words have a strong backwards association but a weak forward association (e.g., fire—blaze). The presence of both words at study but not at test produced overconfident JOLs for these items. As in King, Zechmeister, and Shaughnessy (1980), Koriat and Bjork (2006) demonstrated that study-test practice reduced this bias, but it did not transfer to new items. Explicit training about the foresight bias and the type of item that causes it, however, not only reduced the bias but also transferred to new items.

Retrieval-attempt heuristic. As noted earlier, JOLs often show biased calibration (i.e., overconfidence or underconfidence) and low (but above-chance) relative accuracy. This generalization, however, is limited to immediate judgments of learning—JOLs made immediately after the study of each of the items for which memory is being judged. Almost two decades ago, Nelson and Dunlosky (1991) demonstrated that delaying the making of JOLs from the study of the corresponding item (by interjecting the study and judgment of other items) greatly improved the accuracy of the JOLs. These delayed JOLs showed significantly greater relative accuracy than immediate JOLs and yielded better calibration curves (which compare JOLs to performance over a range of performance levels) than immediate JOLs did. This effect, commonly referred to as the "delayed-JOL effect," has been replicated numerous times since the publication of the original article using different materials and procedures. It has been found to effectively improve the accuracy of JOLs made by children (e.g., Schneider, Vise, Lockl, & Nelson, 2000), older adults (e.g., Connor, Dunlosky, & Hertzog, 1997), and special populations such as patients with schizophrenia (e.g., Bacon, Izaute, & Danion, 2007) and attention-deficit/hyperactivity disorder (e.g., Knouse, Paradise, & Dunlosky, 2006).

Despite the robustness and reliability of the delayed-JOL effect, the cause of the effect has been debated since its discovery. In their original article, Nelson and Dunlosky (1991; see also Nelson & Dunlosky, 1992; Nelson et al., 2004) proposed the Monitoring-Dual-Memories (MDM) explanation for the effect. This explanation centers on the traditional distinction between the short-term and long-term memory systems. According to the MDM hypothesis, to make an accurate JOL one must predict the long-term retention of the stimulus-response association. When an immediate JOL is made for a paired associate, however, the stimulus and response word of the pair being judged are both in short-term memory. The presence of the full pair in short-term memory will therefore interfere with the making of accurate immediate JOLs because they assess both short- and long-term memory. In contrast, a delayed JOL is made after the initial presentation of the word pair has left short-term memory. Delayed JOLs only assess one's long-term memory for an item, which is why they are highly predictive of long-term retention. This also explains why prompting delayed JOLs with the full word pair instead of the stimulus alone can reduce the accuracy of the judgments (Dunlosky & Nelson, 1992).

Spellman and Bjork (1992) countered the MDM explanation for the delayed-JOL effect by suggesting that delaying JOLs improved *memory* for the high JOL items but did not affect metamemory. Their "self-fulfilling prophecy" explanation also assumed that participants attempt to retrieve the response word when making a delayed JOL. Items that are not recalled receive low JOLs while those that are recalled receive high JOLs. Although up to this point this explanation is not entirely different from the MDM explanation, Spellman and Bjork pointed out that items that *were* successfully retrieved therefore received *retrieval practice,* making them more likely to be recalled later. The delayed JOLs create a self-fulfilling prophecy by increasing the likelihood that the successfully retrieved items will also be retrieved later. Kimball and Metcalfe (2003) later provided evidence to support Spellman and Bjork's (1992) explanation for the effect. Specifically, Kimball and Metcalfe showed that re-exposing the response word after a stimulus-prompted delayed JOL was made reduced the relative accuracy of the judgments. This occurred because memory for the low-JOL items was increased by this exposure while memory for the high-JOL items was not affected. Such re-exposure did not affect the relative accuracy or recall of stimulus-prompted immediate JOLs or stimulus-response-prompted delayed JOLs. This evidence provides support for the idea that delayed JOLs are accurate because making such judgments affects memory for the items being judged. The MDM explanation for the delayed-

JOL effect, however, cannot explain the effects of re-exposure without positing an additional mechanism (Kimball & Metcalfe, 2003).

In response to Spellman and Bjork (1992), Nelson and Dunlosky (1992; Nelson et al., 2004) provided evidence to show that delaying JOLs improved the metacognitive accuracy of the judgments, not memory for the items. Specifically, they pointed out that mean recall was the same for items slated with either immediate or delayed JOLs in the Nelson and Dunlosky (1991) dataset, which suggests that delaying JOLs does not boost memory for the items being judged. As Kimball and Metcalfe (2003) pointed out, though, only those items *that were correctly retrieved at the delay* would be expected to get a boost in memorability. Memorability for those items that were *not* correctly recalled at the delay would not be boosted. In contrast, memory for these items might actually be impaired (relative to the immediate-judgment items) because, in the former but not the latter case, they would be rehearsed (if in a massed way). Thus, although overall recall levels in the immediate and delayed JOL cases do not speak to Spellman and Bjork's hypothesis, the recall of retrieved versus unretrieved items does. Using explicit pre-judgment recall attempts, Nelson et al. (2004) showed that delayed JOLs obtain most of their relative accuracy advantage when participants give high JOLs to successfully-retrieved items and low JOLs to items they could not recall. Feedback such as these retrieval attempts can be used to reduce both cognitive and metacognitive errors; results from a non-JOL study by Butler, Karpicke, and Roediger (2008) show that providing students with feedback can increase their recall of low-confidence correct responses, whereas memory for these items will be hampered if feedback is not given..

More recent research by Son and Metcalfe (2005) demonstrated that some low JOLs are made rather quickly—too quickly for retrieval to be attempted. This suggests that JOLs are made in two stages, the first being a quick evaluation of whether one recognizes the cue or not and the second being a target retrieval attempt that can be used to inform the JOL further (see also Metcalfe & Finn, in press). This might help to explain one criticism of the memory-based explanation for the delayed-JOL effect, which is that delayed-JOL items do not consistently show a recall advantage over items for which immediate (or no) JOLs were made. If all delayed JOLs were made based on a retrieval attempt, delayed-JOL items would be expected to consistently show a recall advantage over immediate-JOL items because, as long as target retrieval is successful, of course, the pre-JOL retrieval should enhance performance on the criterion test. Metcalfe and Finn (in press) had participants make either speeded (less than .75 seconds to respond) or unspeeded (unlimited time to respond) delayed JOLs. Memory was enhanced for the unspeeded JOLs relative to the speeded JOLs, presumably because participants in the latter case did not have time to make a retrieval attempt (and instead had to base their JOLs primarily on the familiarity of the cue).

Although the explanation for why the delayed-JOL effect occurs is still under debate, there is consensus on the reliability of the finding itself: empirical evidence has consistently indicated that explicit retrieval attempts help to make delayed JOLs accurate (Finn & Metcalfe, 2007; 2008; Kimball & Metcalfe, 2003; Nelson et al., 2004; Spellman & Bjork, 1992). For this reason, students should be instructed to attempt retrieval and use the outcome as the basis for delayed JOLs. Such retrieval attempts are highly predictive of later test performance (Finn & Metcalfe, 2007; Nelson et al., 2004), which explains why basing delayed JOLs on their outcome predicts test performance so well. As suggested by Thiede (1999), "a new focus for metacognitive training may be to teach students to discriminate between what they know versus what they don't know" (p. 666-667). Having students make delayed JOLs is one way to do this.

Making such delayed retrieval attempts also has the secondary advantage of boosting retention performance by serving as a spaced practice trial (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). It is also advisable to look up the answers to items given low delayed JOLs because corrective feedback will enhance the recall of these items (Butler et al., 2008).

Memory for past test heuristic. Providing students with retrieval opportunities prior to making JOLs increases the accuracy of those JOLs (King et al., 1980; but see Koriat et al., 2002). The effect does not, however, transfer to new items (King et al., 1980). This suggests that although the outcome of a retrieval attempt can be useful for informing JOLs (Finn & Metcalfe, 2007; Nelson et al., 2004), this occurs because it provides information about memory for the item retrieved, *not* information about the retrieval task itself. Recently, Finn and Metcalfe (2007; 2008) demonstrated the Memory for Past Test (MPT) Heuristic. These studies demonstrate that JOLs are based—at least in part—on earlier retrieval outcomes. More specifically, Finn and Metcalfe (2007; 2008) considered whether the MPT heuristic might be responsible for both the decreased calibration and increased relative accuracy associated with the UWP effect (Koriat et al., 2002). They hypothesized that participants were, in part, basing their Trial 2 immediate JOLs on whether or not an item was recalled on Trial 1. Although this contributed to the underconfidence of their Trial 2 JOLs (because some of the items that were incorrect on Trial 1 had been learned on Trial 2), it also contributed to their increased relative accuracy on Trial 2 (because, overall, their Trial 1 recall performance did predict their Trial 2 performance). Although the MPT heuristic can contribute to calibration errors (i.e., underconfidence with practice) it can also increase relative accuracy. Because good relative accuracy is important for the control of study (Thiede, 1999) and there in fact might be some advantage to being slightly underconfident when studying, instructing students to consider their past test performance seems like a simple and effective way to improve the accuracy of their memory monitoring.

Summarization heuristic. Asking learners to summarize what they have read before judging how well they understand it is one way to improve the accuracy of metacomprehension judgments (Thiede & Anderson, 2003). Summarization seems to improve the metacomprehension accuracy of all readers, including high ability readers (Griffin, Wiley, & Thiede, 2008). This improvement, however, is only found when the writing of the summaries and making of the judgments occur at a delay from the reading of the texts (Thiede & Anderson, 2003). Thiede and Anderson (2003) evaluated two potential explanations for this effect. Both involved the participants' use of aspects of the summaries as cues for their judgments, but neither was consistently supported. Despite this earlier conclusion, Griffin et al. (2008) later concluded that summarizing improves metacomprehension accuracy "by increasing the access and salience of cues that are actually predictive of performance that requires inference and situation-model level understanding" (p. 102), but they did not examine this possibility directly. The difference between these two conclusions is that Thiede and Anderson (2003) focused on attributes of the summaries written such as their length without considering how indicative of higher-level understanding such cues are. Future research should attempt to determine what information about the summaries is used as cues when summarizing improves the metacomprehension accuracy of metacomprehension judgments.

Although summarization helps students make more accurate metacomprehension judgments, writing summaries can be a time-consuming endeavor. For this reason, researchers have sought out more time-efficient methods to improve the accuracy of these judgments. One such alternative involves a direct analogue to writing summaries: delayed-keyword generation.

Having students generate keywords (important points from the text) a short time after reading a text but before making a metacomprehension judgment for it makes the judgments more accurate but requires less time than would be required to write a full summary (Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). According to Thiede et al. (2005), delayed-keyword generation improves the accuracy of the judgments by providing the learner with cues that are diagnostic of their understanding of the text. As long as the test will tap the learner's understanding of the text, these cues will also be predictive of performance on that test. Like summarization, keyword generation only improves the accuracy of metacomprehension judgments when employed at a delay from the reading of the texts. Simply delaying metacomprehension judgments, however, does not improve their accuracy (Dunlosky et al., 2005; Maki, 1998a); the delayed judgments must be accompanied by cues that are diagnostic of the learners' understanding of the text (Griffin et al., 2008; Thiede et al., 2005). Summarization and keyword generation are two ways to provide such diagnostic cues.

Knowledge of test heuristic. Letting students know how information will be tested should help them judge their learning more accurately and study more efficiently. Thomas and McDaniel (2007) recently examined how the match between the type of encoding and type of testing affected metacomprehension accuracy. They demonstrated that participants made more accurate metacomprehension judgments when encoding and testing either were both detail-focused or were both concept-focused, which also allowed the students to more effectively control their further study of the materials. These results suggest that if students can be trained to study in a way that is conducive to the type of test they expect, they can improve the accuracy of their metacomprehension judgments while studying. Informing students of the format of the test should also help them to set an optimal "grain size" for their metacomprehension judgments and further increase their accuracy (Dunlosky et al., 2005). Of course, this suggestion also involves teaching students which types of encoding are most appropriate for certain types of test (i.e., increasing their metacognitive knowledge).

Another method for improving the accuracy of metacomprehension judgments is to have students make term-specific metacomprehension judgments (Dunlosky, et al., 2005). In contrast to typical metacomprehension judgments, which are made for an entire text or for large portions of a text, term-specific judgments are made for specific pieces of information from the text pre-selected by the experimenter (or, presumably, the teacher). Rather than judging how well they understand a text, students making term-specific judgments judge their ability to answer specific questions that are very similar to those they will actually be tested with later. These judgments demonstrate greater metacomprehension accuracy than typical metacomprehension judgments, especially when participants are forced to attempt to retrieve answers for each term-specific judgment (Dunlosky et al., 2005) and when feedback is given on the correctness of their answers (Rawson & Dunlosky, 2007). Although this method of improving the accuracy of students' metacomprehension judgments is promising for situations where learners' understanding of text will be tested via retention questions, to date its efficacy has not been evaluated when the test will require a more advanced understanding of the text (i.e., inference or transfer questions). Future research should explore the generality of this method to other question types.

Control

Metacognitive control can be defined as any instance of cognitive control that is informed by metacognitive knowledge or monitoring. Concerning studying specifically, control might involve the choice of which items to study and the allocation of study time (e.g., devoting more study time to certain pieces of information than others) or strategy selection (e.g., choosing what is believed to be an effective way to study certain materials; changing a study strategy when an earlier choice proves inefficient). Regardless of which form(s) of control is used while studying, that control will be informed by the student's metacognitive knowledge and monitoring and will be dependent upon the accuracy of both.

Types of Control

Allocating study time. The general design of laboratory studies that have examined the relationship between monitoring and study-time allocation involves participants studying items or materials for a fixed period of time, judging their learning for those materials, and then re-studying the materials for the amount of time the participants feel is appropriate. In some cases, the mean amount of time spent studying different types of items is compared (e.g., time spent studying easy versus difficult items). Alternatively, a correlation can be calculated between the magnitude of the participants' JOLs and their continued study time. The correlation between JOLs and study time is usually negative (for a review see Son & Metcalfe, 2000), which also obtains for other judgments (i.e., EOLs and FOKs, see Nelson & Leonesio, 1988). In other words, students tend to devote more study time to those items or materials on which they judge their learning to be low for than to those items they believe to be known.

This finding supports the idea that students study to reach a goal state in which to-be-learned information is learned to a pre-set criterion level (Dunlosky & Thiede, 1998). More specifically, the discrepancy-reduction hypothesis forwarded by Dunlosky & Thiede (1998) predicts that students will devote study time to the to-be-learned information until the discrepancy between its JOL and the goal state is eliminated. This account explains the negative correlation typically found between JOLs and study time. One shortcoming of the discrepancy-reduction model, however, is that it does not contain a formal stopping mechanism, or some way for the student to stop studying a piece of information they either feel they are not making progress towards learning or simply cannot ever learn. In other words, the model assumes that continued studying allows students to learn any piece of information to criterion. This is of course unlikely; some information will undoubtedly be too difficult for every student to master and continued studying of the information on its own will likely be in vain (Nelson & Leonesio, 1988). Although Dunlosky and Thiede's (1998) model did not contain a formal mechanism to explain how students decide to cease study, they did suggest one that centered on the student's perceived *rate* of learning. An almost identical mechanism was later proposed and evaluated by Metcalfe and Kornell (2005), who demonstrated that as learners study items over time, their JOLs for the items do not continue to increase. As the rate of perceived learning decreases, students are more likely to cease studying.

Although the negative relationship between JOLs and study time has been demonstrated under varied conditions, neither it nor the discrepancy-reduction hypothesis fully captures the relationship between monitoring and the allocation of study time in all situations. For instance, following Thiede and Dunlosky's (1999) seminal study, Son and Metcalfe (2000) showed that study time is not always allocated preferentially to difficult items; factors such as time pressure and learning goals play a role in the choices made. Participants in Son and Metcalfe's (2000) first experiment were given insufficient time to study all of the items presented, so they devoted study time to the easier items. The discrepancy-reduction model would not predict this pattern of results. When participants in Son and Metcalfe's second experiment studied shorter materials, they devoted more of their time to studying more difficult items. In other words, the pattern of study-time allocation predicted by the discrepancy-reduction model only obtained when there was enough time given for difficult items to be studied.

Thiede and Dunlosky (1999) examined the role of their discrepancy-reduction model in both the allocation of study time *and* study choice (selecting which items to study). More difficult items require more learning to reach a goal-based learning state, so the discrepancy-reduction model predicts that these items should be studied for a longer amount of time than easier items (as in Dunlosky & Thiede, 1998). The model also predicts that more difficult items should also be selected for study *more often* than easier items, but this did not obtain. Instead, Thiede and Dunlosky obtained a positive correlation between participants' JOLs and study choices. In other words, the participants chose to restudy items they assigned higher JOLs to. To account for this result, Thiede and Dunlosky (1999) proposed a multi-level system capable of meeting learning goals in an efficient way. For example, if the learning goal was to learn all items to a criterion level, the original discrepancy-reduction approach would *have* to be used and more difficult items would be studied longer than easier items, resulting in the typical negative correlation between JOLs and study time. If a less-ambitious goal was to be met, such as learning 50% of all items to a criterion, the planning system would choose the easiest 50% of the items to learn, explaining positive correlations between JOLs and study choice in such situations. Within this subset of selected items, though, JOLs and study time would still be negatively correlated as predicted by the original instantiation of the discrepancy-reduction model.

Metcalfe and Kornell (2003) provided further support for the idea that study time is strategically allocated to items. They presented three single-word Spanish-English translations to participants at a time. One of the three translations was easy, one was of medium difficulty, and one was difficult. Participants in their study chose to study the easiest items first before studying the medium-difficulty and difficult items if they had study time remaining. They devoted more *study time* to the medium-difficulty items than to either the easy or difficult items. This pattern of study proved to be efficient; easy items benefited greatly from short amounts of study time, whereas medium-difficulty items required more study time to demonstrate learning gains. More importantly, the pattern of results obtained by Metcalfe and Kornell (2003) supported an alternative to the discrepancy-reduction model of study-time allocation, the region of proximal learning (Metcalfe, 2002). The idea behind the region of proximal learning model is that there is some set of to-be-learned information that is just beyond the current understanding of the learner that will therefore most benefit from further study. This is why both expert and novice adult participants in the studies conducted by Metcalfe (2002) preferentially allocated their study time to items that would most benefit from continued study. Novices (participants who were not experts in Spanish) devoted most of their study time to the easy and medium-difficulty items, whereas experts (participants who declared themselves to be expert Spanish speakers and who scored above 30% on the criterion test) devoted most of their time to the difficult items.

Based on the test performance of these two groups, their strategies were logical. For the novices, the easy and medium-difficulty items benefited from the study time that was allocated to them. Although the experts did not devote much study time to these same items, they nevertheless answered many of them correctly on the test, indicating that they were correct by not devoting study time to them. Novice and expert children in Metcalfe's (2002) studies devoted study time similarly to the adults; novice children devoted most of their study time to easy items while expert children devoted less of their time to the easiest items. Although this suggests that children can be just as strategic in allocating study time as adults, the study-time allocation of the expert children was not as effective as that of the expert adults.

It is important to note that both groups of novices would be considered to have demonstrated inefficient study-time allocation based on the discrepancy-reduction model because they did not devote their study time preferentially to the difficult items. Based on the region of proximal learning idea and the data presented by Metcalfe (2002), though, this strategy proved to be beneficial because allocating study time to the easier items would result in greater overall learning than devoting it to difficult items would. Metcalfe and Kornell (2005) further delineated how the region of proximal learning model affects the allocation of study time by demonstrating that it involves two components: choice (i.e., study choice) and perseverance (i.e., study-time allocation). The first component, choice, has two stages: a yes/no decision to study an item and an evaluation of the priority of the item for study. The decision to study or not is based on an attempt to eliminate items that are already known from the study queue. Priority is determined based on the region of proximal learning; items closest to being learned will be studied first. The second component, perseverance, is based on the stopping mechanism described earlier that considers the rate of learning being experienced for an item. Allocating study time in this way has been demonstrated to be efficient for improving learning and is how adults actually do allocate their study time naturally, both of which support the region of proximal learning model (Kornell & Metcalfe, 2006),

Strategy selection. Sometimes students realize that the way they are currently studying is not producing the desired learning outcomes. In such situations, students might switch (i.e., “toggle”) or modify (i.e., “edit”) their current strategy in favor of a different approach (Winne & Hadwin, 1998; see also Siegler, 2002 for a similar concept). This is another way that metacognition—both knowledge and monitoring—can affect the control of study. In fact, Winne and Hadwin (1998) described editing and toggling as a more advanced form of metacognitive control than self-regulated study. They also cautioned that when a task is complex and metacognitive knowledge about it is low, metacognition could actually hurt performance on the task. In such cases they recommend that the learner be given explicit sub-goals to help scaffold their control. Although young children can make decisions concerning their study choices, they are not as adept at doing so as adults are. For example, Son (2005) recently demonstrated that children could make short-term study choices that optimized their learning, but had difficulty doing so when the decision required thinking about long-term future outcomes (e.g., to space practice instead of massing practice).

Monitoring Accuracy and Control

If metacognitive monitoring is used to control study, it seems logical that the control of study is the most effective when monitoring is accurate.[1] Early studies examining the link between the accuracy of metacognitive monitoring and learning often did not allow students an opportunity to use their monitoring to affect their learning. Typically, then, these studies failed to show a relationship between the two (for a review see Thiede, 1999). It was largely concluded that monitoring accuracy does not affect learning. In contrast, Thiede (1999) allowed students to use their monitoring to affect their learning. Participants studied 36 single-word Swahili-English translations for four seconds each on a computer. After the participants studied the entire list, they made JOLs for each item and then were tested on their memory for the translations. After this test, participants saw all 36 items in an array and could choose to restudy as many of the 36 items as they would like. This procedure was repeated either for six trials or until a participant learned all 36 items. Regression analyses conducted by Thiede (1999) demonstrated that better test performance was related to more accurate monitoring (relative accuracy) and better self-regulated learning (choosing lesser-known items for restudy). Importantly, better test performance was not related to selecting more items for restudy, which further supported the conclusion that more accurate monitoring led to more efficient study.

The accuracy of metacognitive monitoring judgments also affects the self-regulated study of text materials. Participants in a study by Thiede et al. (2003) studied several expository texts and made a metacomprehension judgment for each. The accuracy of their judgments was manipulated between-participants by having the participants generate keywords (important words from the text that capture its main idea) either immediately after reading each text, after reading all of the texts (delayed), or not at all. Participants then answered detail- and inference-based questions for each text and, much as in Thiede (1999), were given an opportunity to select texts for restudy. Participants in the three groups performed equally on the first test, but participants who generated keywords at a delay from the reading of the texts achieved higher relative accuracy than participants in the other two conditions (i.e., a gamma correlation of about +.7 compared to about +.4 for the no keyword group and about +.3 for the immediate keyword group). After studying the selected texts again, participants answered new detail and inference questions as well as the same questions answered earlier. Although participants in the three groups chose to restudy the same number of texts, the groups differed in which texts they chose and how well they performed on the second test. Participants in the delayed-keyword generation group chose to restudy texts that they did not understand well after the first study attempt, whereas participants in the other two groups chose to restudy texts that they understood moderately well. Participants in the delayed-keyword group also performed better than participants in the other two groups on the new questions. As a whole, these results suggest that greater relative accuracy of metacomprehension judgments is associated with better self-regulated learning (choosing to restudy the least-understood texts) and better learning.

The accuracy-control link is apparent in the case of relative accuracy (i.e., Thiede, 1999; Thiede et al., 2003). Poor calibration—in the form of an illusion of knowing (overconfidence) or of not knowing (underconfidence)—has also been demonstrated to have an effect on control processes in that over- and underconfidence can spuriously alter whether or not people choose to study. When students think they have learned the study materials—regardless of whether they really have or not—they will choose not to study; when students think they have not learned the study materials—again, regardless of whether they really have or not—they will choose to study. For example, participants in the first experiment presented by Metcalfe and Finn (2008) studied paired associates on two study-judge-test trials. Half of the items were presented once on Trial 1 and three times on Trial 2 (“1-3” items). The other items were presented three times on Trial 1 and once on Trial 2 (“3-1” items). JOLs were made for all items on both trials. Participants were given the opportunity to select items for restudy after making each Trial 2 JOL but restudy did not actually occur. Although recall was greater for the 3-1 items on Trial 1, it was equated on Trial 2. JOLs for the 3-1 items, however, were greater than those for the 1-3 items on both trials. Most important, participants chose to restudy 1-3 items more often than 3-1 items. This pattern reflected the participants’ misperception that they knew the 1-3 items less well than they knew the 3-1 items. This effect obtained even though the relative accuracy of the Trial 2 JOLs was equal for both subsets of items. This suggests that distortions of calibration might exert a negative effect on the control of study even in situations where relative accuracy is good (i.e., Koriat et al., 2002; Metcalfe & Finn, 2008). This possibility deserves further investigation, as little research has examined the effect of poor calibration on study efficiency or study choices.

It is not enough for students to make accurate metacognitive judgments—they also need to know how to transform them into a study plan. One such plan is suggested by Metcalfe’s (2002) Region of Proximal Learning (RPL) model, which suggests that students should allocate study time to the information that is most amenable to learning (Metcalfe, 2002; Metcalfe &

Kornell, 2003). As demonstrated by Kornell and Metcalfe (2006), this strategy is effective for enhancing learning, so students should be trained to allocate their study time in this way. As suggested by Metcalfe and Kornell (2005), the first step should be for students to eliminate from study information that they already know. Making accurate metacognitive judgments here will be critical, as students might mistakenly eliminate information that they do not yet know *well;* caution should be exercised when making such decisions. It would be better to restudy known information than to not at all study information that could actually benefit from further study. Once already-known information is eliminated from study, students should determine how well they have already learned the remaining to-be-learned information and devote study time to that information which is most amenable to learning—e.g., to information or items given the highest JOLs. When that information is learned well, it too should be eliminated from study so more difficult information can be attended to. Studying in this way should produce the largest learning gains in a limited amount of time regardless of the ultimate learning goal (i.e., to learn all materials to mastery or to just learn some portion of the information). Loftier goals will likely require more overall study time and better time management, so students might also need to be trained on how to set goals and how to manage their time. Training on goal setting and time-management has been shown to improve self-regulated learning behaviors when combined with training on metacognition (Azevedo & Cromley, 2004).

Of course, there is a difference between information that has, for all purposes, been committed to memory (i.e., that can readily and reliably be retrieved from memory) and newly-learned information that might be recalled once but might not be recalled again in the future. As demonstrated by Kornell and Bjork (2008), good students naturally "drop" items from study when using flashcards, but often do so in ineffective ways. For example, in their experiments some students dropped items that seemed learned (i.e., because they were successfully recalled once) that in fact were not *well* learned. This becomes increasingly common as more items are dropped and the intervals between study trials for individual items decrease (see Pyc & Rawson, 2007 for a demonstration of the effectiveness of dropping items from study and its dependency on study intervals). Other students in Kornell and Bjork's (2008) experiments dropped items that they felt would be too difficult to learn in the given time constraints, which hurt their overall performance because the items might have been learnable if they had studied the items repeatedly. These results suggest that before students should be instructed to "drop" items from study, they must be taught when it is advantageous to drop items and when it is not.

Summary and Future Directions

In this chapter we reviewed metacognition, its role in studying, and how errors in monitoring can reduce the efficacy of study. We also reviewed methods that improve the accuracy of metamemory and metacomprehension judgments and suggested ways for students and teachers to implement them in the classroom and during study. Although we based these suggestions on empirical evidence, further research is needed to ensure that implementing these methods as we suggest is effective. Research should also continue to find even more effective ways of improving metacognitive knowledge, monitoring, and control.

References

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96,* 523-535.

Bacon, E., Izaute, M., & Danion, J. (2007). Preserved memory monitoring but impaired memory control during episodic encoding in patients with schizophrenia. *Journal of the International Neuropsychological Society, 13,* 219-227.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55-68.

Bjorklund, D. F., & Green, B. L. (1992) . The adaptive nature of cognitive immaturity. *American Psychologist, 47,* 46-54.

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2008). Correcting a meta-cognitive error: Feedback enhances retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918-928.

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

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374-380.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate monitoring and accessibility hypotheses. *Journal of Memory and Language, 52,* 551-565.

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning [Special issue]. *Journal of General Psychology, 132,* 335-346.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98,* 37-56.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition, 36,* 813-821. Finn, B., & Metcalfe, J. (2007). The Role of Memory for Past Test in the Underconfidence with Practice Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 238-244.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 238-244.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58,* 19-34.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288-299.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906-911.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119-136.

Glenberg, A. M., Wilkinson, A, C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597-602.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36,* 93-103.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6,* 685-691.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126-135.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 93,* 329-343.

Kimball, D. R, & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31,* 918-929.

Knouse, L. E., Paradise, M. J., & Dunlosky, J. (2006). Does ADHD in Adults Affect the Relative Accuracy of Metamemory Judgments? *Journal of Attention Disorders, 10,* 160-170.

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.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36,* 416-428.

Koriat, A., & Bjork, R. A. (2006). Mending Metacognitive Illusions: A Comparison of Mnemonic-Based and Theory-Based Procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133-1145.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131,* 147-162.

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.

Kornell, N., & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits—and costs—of dropping flashcards. *Memory, 16,* 125-136.

Maki, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26,* 959-964.

Maki, R. H. (1998b). Test predictions over text material. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349-363.

Metcalfe, J., & Finn, B. (2008). Evidence that Judgments of Learning are Directly Related to Study Choice. *Psychonomic Bulletin Review, 15,* 174-179.

Metcalfe, J., & Finn, B. (in press). Familiarity and Retrieval Processes in Delayed Judgments of Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Metcalfe, J., & Kornell, N. (2003). The Dynamics of Learning and Allocation of Study Time to a Region of Proximal Learning. *Journal of Experimental Psychology: General, 132,* 530-542.

Metcalfe, J., & Kornell, N. (2005). A Region of Proximal Learning Model of Study Time Allocation. *Journal of Memory and Language, 52,* 463-477.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851-864.

Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology, 10,* 257-260.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2,* 267-270.

Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3,* 317-318.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676-686.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: The MIT Press.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A Revised Methodology for Research on Metamemory: Pre-judgment Recall And Monitoring (PRAM). *Psychological Methods, 9,* 53-69.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & 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. Retrieved from http://ncer.ed.gov.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35,* 1917-1927.

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

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435-451.

Schneider, W., Vise, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning (JOL) task. *Cognitive Development, 15,* 115-134.

Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31,* 1258-1266.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), Microdevelopment: Transition processes in development and learning (pp. 31-58). Cambridge, UK; Cambridge University Press.

Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies [Special issue]. *Journal of General Psychology, 132,* 347-363.

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.

Son, L. K., & Metcalfe, J. (2005). Judgments of Learning: Evidence for a Two-Stage Model. *Memory & Cognition, 33*, 1116-1129.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3,* 315-316.

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. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28,* 129-160.

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

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024-1037.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1267-1280.

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition, 35,* 668-678.

Vesonder, G., & Voss, J. (1985). On the ability to predict one's own responses while learning, *Journal of Memory and Language, 24,* 363-376.

Weaver, C. A., III, Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse Comprehension: Essays in Honor of Walter Kintsch* (pp. 177-193). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Weaver, C. A., III, & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science, 8,* 318-321.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16,* 3-118.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

Winne, P. H. (1995). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8,* 327-353.

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

Author Note

Michael J. Serra, Department of Psychology, Columbia University. Janet Metcalfe, Department of Psychology, Columbia University.

This research was supported by National Institute of Mental Health grant RO1-MH60637. We thank John Dunlosky for his help and comments.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, 406 Schermerhorn Hall, 1190 Amsterdam Avenue, New York, NY 10027. E-mail: jm348@columbia.edu

Footnote

[1]Although we have stressed making *accurate* metacognitive judgments in this chapter, there might be situations where it is advantageous to make inaccurate judgments. In particular, there might be advantages to being underconfident when making learning judgments. Underconfident judgments should result in extra study, and might be appropriate in real life studying situations where the test is typically delayed from the study episode (as opposed to experimental situations where the test typically occurs shortly after the study episode). It is difficult to imagine situations in which making judgments with poor relative accuracy or overconfident judgments might be advantageous for adults (young or old). For children, however, it has been argued (e.g., Bjorklund & Green, 1992) that overconfidence can be advantageous because their cognitive abilities and knowledge are still developing. If children accurately knew that they would perform poorly on most things they attempted to do (i.e., if their self-efficacy matched their abilities), they might not attempt to do those things, and therefore might never learn how to do them.