Oxford Handbooks Online

Metamemory and Education

Nicholas C. Soderstrom, Carole L. Yue, and Elizabeth Ligon Bjork

The Oxford Handbook of Metamemory *(Forthcoming)*
Edited by John Dunlosky and Sarah Tauber

Subject: Psychology, Cognitive Psychology, Cognitive Neuroscience
Online Publication Date: Feb 2015
DOI: 10.1093/oxfordhb/9780199336746.013.6

[-] Abstract and Keywords

Metamemory refers to one's knowledge and understanding of how learning and memory operate, as well as the interplay between the monitoring and controlling of one's own memory and learning. This chapter reviews and evaluates the current state of metamemory research—basic, applied, and survey—with respect to its educational implications. Among the relevant aspects of metamemory discussed are the growing number of findings that, although people's beliefs and ongoing assessments of their own learning can be sometimes quite accurate, they can also be very much at odds with their actual learning and create illusions of competence that can lead to the adoption of ineffective learning strategies. The chapter gives special consideration to procedures that might help foster metamemorial sophistication, with the goal of optimizing self-regulated learning. It concludes with some general remarks regarding the educational implications of metamemory research and offers some promising directions for future research.

Keywords: metamemory, education, memory, learning, students

Introduction

For instructors and students alike, academic achievement—from kindergarten to post-baccalaureate training programs—has arguably never been more consequential or competitive than it is today. Success in the classroom impacts, among other things, educational placements, college admissions, job prospects, and promotions. It is thus imperative to elucidate those factors that bear on academic achievement.

Although there are, undoubtedly, myriad factors implicated in classroom success, including social and motivational influences, the focus of this chapter is on one crucial factor: metamemory. Broadly considered, *metamemory* is a type of metacognition that refers to one's knowledge and understanding of how learning and memory operate, as well as the interplay between the monitoring and controlling of one's own memory and learning. To what extent are instructors and students aware of what is beneficial for long-lasting learning? How do students know when to stop studying for an upcoming exam? What steps should learners take to remedy memory failures? What measures can be taken to optimize self-regulated learning? Questions such as these have driven decades of metamemory research, out of which has emerged a wealth of findings that are particularly relevant for learning within educational contexts.

In this chapter, we review and evaluate the current state of metamemory research with respect to its educational implications. We discuss research—basic, applied, and survey—that has examined what people believe about learning and the extent to which learners can accurately assess and regulate their own learning. The research we review is not meant to be exhaustive, but rather selective and exemplary. To foreshadow, people's beliefs and ongoing assessments of their own learning can, under some circumstances, be quite accurate, but they can also be vastly misaligned with actual learning, creating illusions of competence. Therefore, we have given special consideration to procedures that might help foster metamemorial sophistication with the goal of optimizing self-regulated learning. Finally, we conclude with some general remarks regarding the educational implications of metamemory research and offer our thoughts on some promising directions for future research in this area.

Surveys of Students' Beliefs About Learning

Effective learning depends, in large part, on the knowledge, understanding, and implementation of effective learning strategies. Therefore, much metamemory research has focused on what strategies learners believe support effective learning, how often such strategies are used, and the degree to which these strategies correspond to what previous research has identified as being objectively effective for learning. In this section, we selectively summarize survey research that has illuminated important findings regarding learners' beliefs about their own learning and the strategies students use while studying.

Perhaps the most straightforward method of assessing learners' beliefs about their own learning is simply to ask them to identify the study strategies that they regard as most effective. To our knowledge, one of the earliest studies to engage in a large-scale survey of student study strategies is one by Mackenzie (1994). Although Mackenzie's primary purpose was to examine the relationship between test preparation and anxiety, she used written self-reports and student interviews to generate a 29-item questionnaire about study strategies for subsequent research. One of the most commonly reported strategies was rereading, with students reporting rereading class assignments "most of the time." Although students also reported trying to anticipate and prepare for future exam questions "fairly often," that strategy may or may not have involved self-testing, which is a normatively effective learning strategy (see Roediger & Karpicke, 2006a). It is conceivable that after deciding on a possible question, students tried to respond as though they were actually taking a test. However, it is also possible (and perhaps more likely, given that many of the self-reported strategies involved some level of rereading) that after coming up with a possible question, students reread material related to that question rather than trying to answer it from memory. Indeed, one of the less frequent strategies students reported was to "practice writing answers under exam conditions," an approach that could be seen as reflecting an understanding of the benefits of retrieval practice as a study strategy (see Roediger & Karpicke, 2006a, for an excellent review of the evidence for the benefits of retrieval practice as a learning strategy).

More recent survey work on students' study strategies has incorporated evidence from empirical research to assess how students' beliefs align (or do not align) with research findings. McCabe (2011), for example, investigated undergraduates' awareness of six empirically supported learning strategies: (a) presenting material in multiple sensory modalities (e.g., auditory and verbal) rather than a single modality (Mayer & Moreno, 1998), (b) presenting material in a static media situation rather than an animated one (Mayer, Hegarty, Mayer, & Campbell, 2005), (c) including low-interest extraneous details rather than those of high-interest (Mayer, Griffith, Jurkowitz, & Rothman, 2008), (d) testing rather than restudying (for a review, see Roediger & Karpicke, 2006a), (e) temporally spacing study episodes rather than massing them (for a review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), and (f) generating to-be-remembered information rather than simply reading it (for a review, see Bertsch, Pesta, Wiscott, & McDaniel, 2007). Were students aware of these learning principles? The answer is a resounding, "no." For five of the six learning strategies, students overwhelmingly endorsed the less effective version, a pattern that was especially evident in the spaced-versus-massed scenario in which only 6.67% of students accurately endorsed the effectiveness of spaced study over massed study. What makes this finding particularly striking is that the spacing effect is, arguably, the most robust effect in all of memory research, its original empirical demonstration dating back to over a century ago (Ebbinghaus, 1885/1964)! The one effective strategy that students seemed to appreciate, albeit weakly, was the generation effect—about 50% accurately endorsed generation as a beneficial strategy.

In a follow-up study, McCabe (2011) investigated whether explicit instruction on applied learning and memory topics increased students' awareness of effective learning strategies. To do so, the previously used survey was administered to groups of students with varying educational experiences with applied learning and memory topics. Consistent with the notion that educational intervention may improve metacognitive awareness of effective learning strategies, students who had taken an introductory psychology or cognition course were more likely to endorse the more effective strategies compared to students that had not taken such courses. Furthermore, students who received targeted instruction on the six effective learning strategies included on the survey showed the highest levels of correct endorsements. Thus, although the strong tendency of noninstructed students to endorse or prefer ineffective strategies to more effective ones is troublesome, it is reassuring that educational interventions may help in correcting these misperceptions. The degree to which instruction directly influences the learner's implementation or adoption of effective learning strategies, however, is still an open question.

In addition to direct instruction, it is possible that students are able to learn effective study strategies through experience. In a classroom survey of students' study habits, Carrier (2003) asked students to self-report their study strategies after two class exams. Students received their scores on the first exam before taking the second exam, so it is plausible that they would have modified any ineffective strategies employed while studying for the first exam to improve their performance on the second exam. Unfortunately, despite the slight (nonsignificant) negative correlation between rereading and score on the first exam, approximately the same proportion of students reported using rereading as a study strategy for the second exam (.65 and .64, respectively). Other popular strategies, such as recopying notes or outlining chapters, had negative or no correlations with exam scores but were used with equal frequency to study for both exams. On the other hand, several strategies that showed a positive correlation with performance on the first exam (e.g., taking chapter notes, highlighting the textbook) were used more often as study strategies for the second exam. This pattern may indicate that students are, to some degree, able to learn from experience and subsequently implement more effective strategies. It is also possible, however, that students may have adjusted their study strategies owing to the differing nature of the exams: the first exam was closed book, while the second was open book. Thus, it may be that students' perceptions of which strategies are most helpful vary by the expected type of exam, as well as recent experience.

Educators often assume (or hope) that students select study strategies and schedules based on what would help them learn the most, or at least what would optimize their performance on a given assessment; however, other factors (e.g., time, interest) may interfere with students' pursuit of this goal. To illustrate, in a survey of 472 college students investigating not only what strategies students choose to use during their own study, but also why they do so, Kornell and Bjork (2007) found that most students choose to study "whatever is due soonest," with the next most popular option being to study whatever subject "they feel they are doing the worst in." In addition, students' responses demonstrated an apparent lack of awareness of forgetting: 64% of students reported not restudying material once they felt like they understood it, while only 36% of students reported that they would go back to study or test themselves later on that piece of information. These results are in line with other findings on the stability bias; that is, the tendency for individuals to overestimate how much they will recall and to underestimate how much they will forget over time (see Kornell & Bjork, 2009), an issue that we return to later in this chapter.

In terms of specific study strategies, Kornell and Bjork (2007) found that most students reported rereading sections of a text that they had previously marked in some way. Interestingly, although about 90% of students reported using self-testing as a study technique, most of them reported doing so in order to find out how well they had learned the material, rather than because they believed that self-testing helped them learn. This lack of understanding of one of the main benefits of testing (i.e., its power as a tool for learning) may prevent some students from using it effectively (i.e., truly attempting to retrieve information from memory, rather than just reading practice test questions), or it may prevent the other 10% from using testing as a study strategy at all.

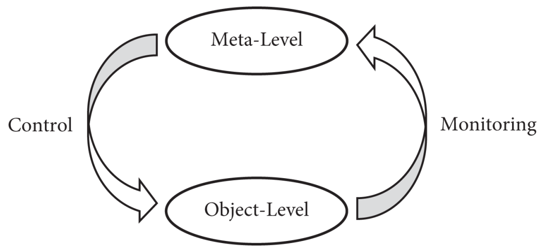
The findings of Kornell and Bjork (2007) paint an interesting picture of the studying behavior of a large sample of undergraduates and the factors affecting their study strategy selections. They do not, however, let us know anything about how the different strategies students report using relate to actual performance in the classroom. Although we can be confident that certain strategies (e.g., trying to retrieve information versus rereading it and spacing versus massing study opportunities) produce better learning in the laboratory, whether they also produce better learning in the classroom is essential to know as well. After all, if study strategies tested in the lab are not effective in the classroom or in real-world learning, then we have no reason to encourage students to use those strategies. To address this issue, Hartwig and Dunlosky (2012) examined the relationship between the study strategies students reported using and their grade point averages (GPA). The first part of their survey included the same questions that Kornell and Bjork (2007) had asked, and they obtained strikingly parallel results. As found by Kornell and Bjork, most students reported using self-testing primarily to evaluate their progress, rather than because of a belief that testing itself promotes learning. A series of nonparametric correlations revealed that the reported use of testing with practice questions or problems was positively related to GPA, while the reported use of flashcards, which are sometimes considered self-testing tools, was not related to GPA. Hartwig and Dunlosky also found that students with higher GPAs were more likely to report scheduling their study time than did students with lower GPAs, and that low performers were more likely to study whatever is due soonest. Interestingly, there was no significant benefit of spacing in terms of GPA, although the trend was in the expected direction (i.e., students who spaced their study had numerically higher GPAs than those who did not). Thus, in general, high performers tended to utilize better study strategies.

Homing in specifically on students' awareness of testing as a study strategy, Karpicke, Butler, and Roediger (2009) conducted a survey of 177 undergraduates to determine whether students' self-reports about their study behavior matched laboratory data indicating that students are generally unaware of the benefits of retrieval practice compared to rereading as a study strategy. They asked students to write down and rank any study strategies they used. By far, rereading was the most popular strategy, with 84% of students listing it as one of their strategies of choice and 55% ranking it as their number one strategy. Only 11% of students reported practicing retrieval at all. In a second question, students were asked to make an explicit decision between rereading, testing, or some other study technique. When testing involved not being able to restudy the material afterward, less than 20% of students chose testing, and only about half of those stated that they believed testing would help them perform well on a future exam. Even when testing was presented with the possibility of subsequent restudy, less than half the students opted for testing, and even fewer (3%) said they would do so because practicing retrieval would help them on a later test. Instead, most said they would use the feedback to guide future studying. Karpicke et al. offered the suggestion that the temptation of rereading, rather than testing, is due to the illusion of competence/sense of fluency that comes from repeated studying. Similar to Kornell and Bjork's (2007) findings, these results indicate that even when students do choose testing, they do not do so for the entirely "right" reasons.

Thus far, we have primarily focused on students' beliefs regarding study strategies that are clearly effective (e.g., testing) or clearly ineffective (e.g., rereading); however, students also report using strategies for which the empirical evidence is ambiguous. Highlighting, for example, is a strategy that many students employ either during initial reading or subsequent study (e.g., Carrier, 2003; Kornell & Bjork, 2007; Bell & Limber, 2010). Unfortunately, the research on highlighting has obtained mixed results: some studies have shown a benefit (e.g., Fowler & Barker, 1974; Yue, Storm, Kornell, & Bjork, in press), whereas others have not (e.g., Peterson, 1992; Wade & Trathen, 1989). A variety of factors seem to be involved, most notably the student's ability to distinguish relevant from irrelevant information (Bell & Limber, 2010; Stordahl & Christensen, 1956). Given the prevalence of highlighting in students' self-reports of study strategies, it may be advisable to train students to use other techniques in combination with highlighting, as suggested by Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013).

In this section, we have described survey research illuminating students' beliefs about how learning and memory operate, as well as the types of strategies that students use when studying. By and large, students endorse and utilize relatively ineffective learning strategies—sobering news to educators, to be sure. Later, we discuss potential ways to foster students' metamemorial sophistication with the goal of optimizing self-regulated learning. Next, however, we discuss the degree to which learners can accurately monitor and control their own ongoing learning.

Metamemory Monitoring and Control



Click to view larger

Figure 1 . Nelson and Narens's basic metacognitive framework illustrating the bidirectional flow of information between the object-level (cognition) and the meta-level (metacognition).

(Adapted from Nelson & Narens, 1990.)

In addition to denoting one's knowledge and understanding of how learning and memory operate, metamemory also encompasses the interplay between the monitoring and controlling of one's ongoing learning. Over two decades ago, Nelson and Narens (1990) proposed what is still regarded as the dominant conceptual framework for understanding this interplay, defining *monitoring* as the subjective evaluation of one's own memory processes and *control* as the resulting cognitive and behavioral consequences of such evaluations. As argued by Nelson and Narens and illustrated in Figure 1, the relationship between monitoring and control can be conceptualized in terms of object-level and metalevel cognitions. Object-level cognitions involve basic operations such as encoding and retrieving information, whereas metalevel cognitions refer to those that oversee the object-level, updating object-level operations when necessary. Stated in their own words, "the meta-level *modifies* the object level. In particular, the information flowing from the meta-level to the object-level either changes the state of the object-level processes or changes the object-level process itself" (Nelson & Narens, 1990, p. 127, italics original).

To illustrate the interplay between the object- and meta-level, consider a student studying for an upcoming exam. The student might realize after reading a chapter in a textbook that he or she did not fully understand the information. At this point, the meta-level is monitoring the object-level. To rectify the situation, the student might choose to reread the material and take better notes, behavioral modifications that were initiated by monitoring processes at the meta-level. Notice that inherent in this example is the notion that monitoring plays a causal role in control decisions—that is, subjective experiences can directly shape behavior. Indeed, Nelson and Narens (1990) made such an assumption explicit, stating, "*A system that monitors itself (even imperfectly) may use its own introspections as input to alter the system's behavior*" (p. 128, italics original). As we discuss later, recent metamemory research (e.g., Metcalfe & Finn, 2008; Rhodes & Castel, 2009; Soderstrom & Rhodes, 2014) has validated the monitoring-affects-control assumption, bolstering the importance of conducting metamemory research and underscoring the crucial role that metamemorial processes play in learning.

According to Nelson and Narens (1990), the bidirectional flow of information between the object-level and metalevel operates during all three major stages of learning: acquisition, retention, and retrieval. Essentially, effective learning requires a continual interplay between accurate monitoring and the necessary control decisions in response to that monitoring throughout the entirety of the learning process. Hence, for effective learning to occur, the accuracy of monitoring is paramount—after all, if individuals are unaware of what they have and have not learned, it is unreasonable to expect that they will take the remedial steps necessary to improve their own learning. In short, effective control is predicated on effective monitoring. We now discuss laboratory- and classroom-based research that has investigated the bases of subjective memory assessments, the validity of such assessments, and the degree to which they influence subsequent behavior.

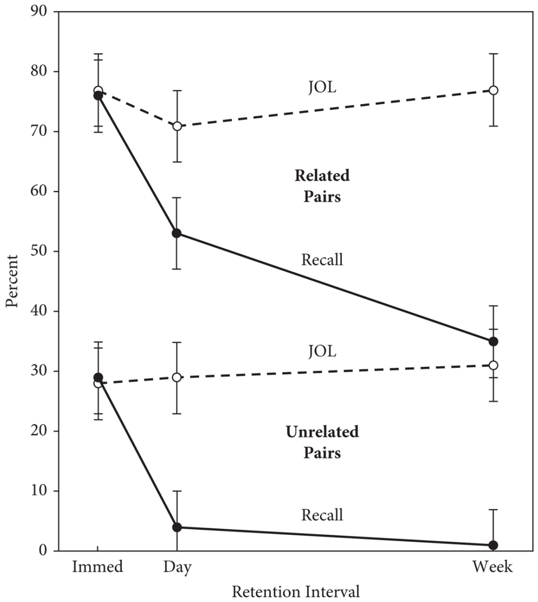
Monitoring One's Own Memory: Judgments of Learning

Although a wide variety of methods have been utilized to examine how, and to what degree of accuracy, people can monitor their own learning and memory (see Part 3 of the present volume), the most common contemporary approach is to ask people to predict their own future memory performance. For this reason, and because students likely implicitly make such assessments in their own lives—say, when deciding whether they have adequately studied certain material for a future exam—we constrain our discussion in this section to cases where learners are asked to forecast their own memorial experiences. Research in this vein has largely focused on eliciting judgments of learning (JOLs), in which people are asked during acquisition, or the encoding phase, to assess the likelihood of remembering information on a later test. As a simple illustration, a learner might study Spanish-English vocabulary pairs such as *Olvidar—Forget*, and then be asked, "On a scale from 0% to 100%, what is the likelihood that you will remember the English translation (*Forget*) if presented with its Spanish counterpart (*Olvidar*)?" JOLs are typically made after each item in a study list, although they can also be solicited once at the beginning of a list (called an aggregate JOL) in which a global prediction is made for all of the studied items (e.g., "How many of the 30 vocabulary pairs do you think you will remember on a later test?"). This relatively simple empirical approach permits an examination of how learners conclude (decide) that material has, or has not, been learned and how well such predictions correspond with actual performance on a subsequent test.

Out of decades of basic metamemory research have emerged two dominant theoretical accounts that explain how learners approach judging the likelihood that information will be remembered later. The *direct-access account* argues that people directly access memory trace strength for each item during study and make their judgments accordingly (e.g., Hart, 1967; Jang & Nelson, 2005). If the memory trace is perceived to be strong, a relatively high

JOL will be given. Conversely, if the trace is weak, the JOL given will be relatively low. As an alternative view, Koriat (1997) proposed the *cue-utilization account*, which suggests that JOLs are based on a variety of cues or heuristics, some of which may be more valid in predicting memory performance than others—that is, judgments are inferential in nature.

One prediction made by the direct access account is that JOLs should always parallel memory performance, because both JOLs and memory performance are based on trace strength; indeed, early work showed that JOLs matched memory performance relatively well (see, e.g., Arbuckle & Cuddy, 1969). More recent metamemory literature, however, is rife with examples in which predictions are not diagnostic of future performance (e.g., Benjamin, Bjork & Schwartz, 1998; Castel, McCabe, & Roediger, 2007; Koriat & Bjork, 2005; Koriat & Goldsmith, 1996; Koriat, Bjork, Sheffer, & Bar, 2004; Mazzoni & Nelson, 1995; Rhodes & Castel, 2008; Soderstrom & McCabe, 2011), revealing interesting illusions of competence and providing compelling evidence in favor of the inferential view of JOLs. We next review some common misconceptions among learners as indexed by their JOLs (for a more comprehensive review, see Schwartz & Efklides, 2012).



Click to view larger

Figure 2 . Mean judgment of learning (JOL) and recall as a function of retention interval (immediate, 1 day, or 1 week) for related and unrelated word pairs.

(Adapted from Koriat et al., 2004.)

In general, learners tend to be overconfident in their memory abilities and frequently demonstrate what Kornell and Bjork (2009) have termed the *stability bias*, which refers to the finding that people often discount factors that lead to impaired or enhanced learning, holding instead to the view that performance will remain stable across time (see also Kornell, 2011). In a particularly striking example of a stability bias, Koriat, Bjork, Sheffer, and Bar (2004) had participants make item-by-item JOLs regarding the likelihood that they would remember word pairs on an immediate test, after one day, or after one week (manipulated between subjects). As illustrated in Figure 2, participants produced a pattern of results demonstrating apparent insensitivity to retention interval, giving equivalent JOLs across the three retention intervals; whereas as expected, recall performance decreased as a function of retention interval. At the same time, the JOLs given to these word pairs indicated a strong sensitivity to the relatedness of the word pairs. This finding led the authors to conclude that encoding fluency, or how easily information is processed during study, can largely drive JOLs (see also, Begg, Duft, Lalonde, Melnick & Sanvito, 1989; Koriat, 2008; Undorf & Erdfelder, 2011), even at the expense of other extremely relevant information—in this case, retention interval. Only when retention interval was manipulated within-subjects did participants appear to take retention interval into account when making their JOLs. Other work supports the conjecture that *retrieval* fluency (i.e., how easily to-be-remembered information is retrieved during learning), like encoding fluency, can also have a dramatic impact on JOL magnitude (see, e.g., Benjamin et al., 1998; Hertzog, Dunlosky, Robinson, & Kidder, 2003). These results suggest that some commonly used study behaviors that increase fluency (e.g., rereading) may produce illusions of knowing whereby students erroneously interpret fluency as being a reliable index of long-term retention.

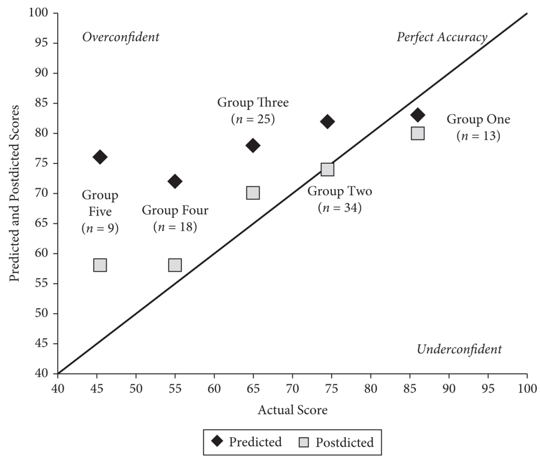
Consistent with the previously discussed survey research, learners' JOLs also appear to underestimate the benefits of empirically validated learning strategies. Zechmeister and Shaughnessy (1980), for example, showed that people give higher JOLs for words that are studied under massed conditions relative to words that are studied in a temporally spaced fashion, even though spacing is more effective for long-term learning than massing (i.e., the spacing effect; see Cepeda et al., 2006). Such a bias to endorse massed study is even evident *after* learners experience the value of spacing (Kornell & Bjork, 2008), and it is now clear that learners, when permitted to choose their own study schedule, favor massing (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013). Similarly, Roediger and Karpicke (2006b) demonstrated that learners also underestimate the mnemonic benefit of retrieval practice—specifically, participants mistakenly predict better long-term learning of prose passages when the passages are restudied multiple times compared to when the passages are tested multiple times. Actual learning, however, showed the opposite pattern: Long-term retention increased as a function of the amount of testing opportunities during acquisition, showcasing that retrieval practice strengthens the retrieved information, rendering it more likely to be remembered in the future (i.e., the testing effect; see Roediger & Karpicke, 2006a). Given that repeated study can produce a sense of fluency during acquisition and that it often outperforms spaced study and testing on the short-term, this pattern of results can also be couched in terms of a stability bias, with leaners ostensibly thinking, "If it's helping me now, it will help me later."

Learners can also be misled by irrelevant perceptual information. Rhodes and Castel (2008), for example, showed that participants give higher JOLs for words presented for study in a large font relative to words presented in a small font, whereas future memory performance does not differ as a function of font size. Of current debate is the underlying cause of this result: namely, whether font-size effects on JOLs are driven by perceptual fluency, beliefs about memory, or both (see Mueller, Dunlosky, Tauber, & Rhodes, 2014). Analogous results have been demonstrated for blurry versus clearly presented words (i.e., higher JOLs are given for clear words despite no differences in later memory; Yue, Castel, & Bjork, 2013) and for auditory information (i.e., loudly spoken words are erroneously predicted to be more memorable than quietly spoken words; Rhodes & Castel, 2009; Soderstrom & Rhodes, 2014). Thus, these studies suggest that manipulating perceptual characteristics of to-be-learned information has no bearing on actual memory, although learners think it does. We note, however, that some work has indicated that presenting information in a perceptually disfluent manner can, under some circumstances, enhance memory, presumably because it promotes deeper, more elaborate encoding processes (see, e.g., Diemand-Yauman, Oppenheimer, & Vaughn, 2010; Sungkhasette, Friedman, & Castel, 2011). Knowledge of such results is important for instructors, who must make decisions about how best to construct their lecture presentations, and for students, who frequently encounter perceptually variable text (e.g., bolded words, large subheadings) while studying.

The last several examples have highlighted cases in which JOLs and later performance diverge; fortunately, there are methods of improving JOL accuracy. If permitted practice with multiple study-test phases, for example, the accuracy of participants' memory predictions often improves markedly (e.g., Benjamin, 2003; Finn & Metcalfe, 2007; Castel, 2008; Koriat, Sheffer, & Ma'ayan, 2002), presumably because prior testing equips the learner with information that is diagnostic of future recall—namely, whether items were previously recalled or not. In addition, making JOLs after a delay—that is, by inserting time between studying material and making predictions for that material—significantly improves their accuracy, a finding termed the *delayed-JOL effect* (Nelson & Dunlosky, 1991; for a review, see Rhodes & Tauber, 2011). The leading candidate explanation for this effect is that delayed JOLs, unlike immediate JOLs, are not contaminated by information from short-term memory (e.g., encoding fluency). Rather, delayed JOLs are based on information from long-term memory, such as the retrievability of the to-be-learned material. Because the later memory test is also based on information from long-term memory (i.e., retrievability), delayed JOLs tend to be quite accurate. This hypothesis has been corroborated by work showing that the delayed-JOL effect is limited to situations in which only the cue word in a pair (e.g., *dog -?* for the pair *dog —spoon*) is presented for the delayed JOL; delaying the judgment is not helpful when the judgment is made in the presence of the cue and target (Connor, Dunlosky, & Hertzog, 1997; Dunlsoky & Nelson, 1992). Active retrieval of the target word may also produce a sort of self-fulfilling prophecy whereby the relationship between JOLs and later recall is bolstered because successful retrieval at the time of the JOL increases the likelihood of successful retrieval on the later test (see Spellman & Bjork, 1992). One clear educational implication of this work on delaying JOLs is that students should be encouraged to test themselves as a method of self-assessment.

Thus far, we have discussed JOL accuracy in the context of laboratory experiments, but how well do learners predict their own performance in more naturalistic learning contexts? There are, of course, potentially important differences between the hyper-controlled laboratory and more naturalistic learning environments that may dampen the enthusiasm of researchers and educators alike when it comes to generalizing laboratory findings to real-world situations (for a more in-depth discussion on this issue, see Maki & McGuire, 2002). Two often encountered criticisms of laboratory research from the perspective of classroom instructors and educational scientists are that the materials are relatively simple and that participants may not be particularly motivated to perform given the inconsequential and obscure nature of laboratory-based tasks. Fortunately, numerous studies, including a host focused on JOL accuracy, have been conducted using more educationally relevant materials and/or in classroom settings. For example, paralleling JOL research using simpler materials, Glenberg and Epstein (1985) developed an experimental paradigm in which participants read and then rate their comprehension of multiple text passages before taking a test on those passages. This area of research, called *metacomprehension*, has also revealed that learners typically exhibit overconfidence and poor predictive accuracy (for a review, see Maki, Shields, Wheeler, & Zacchilli, 2005).

The general finding of overconfidence in such tasks has been the impetus for several training studies in metacomprehension. Koch (2001), for example, demonstrated that participants who engaged in ranking their comprehension abilities and disabilities in a hierarchical fashion showed greater benefits to understanding physics texts than did participants in a control condition in which these self-reflections were not required. Likewise, Thiede and colleagues have shown that summarizing to-be-learned texts after a delay (Thiede & Anderson, 2003) or generating keywords pertaining to those texts (Thiede, Dunlosky, Griffin, & Wiley, 2005) can improve metacomprehension accuracy. Delaying the generation of summaries and keywords, it is argued, helps the learner focus on important information, to build relationships among concepts, and encourages self-testing. Thus, these training studies illustrate several ways in which the gap between subjective learning and objective learning of text materials might be bridged.



Click to view larger

Figure 3 . Mean predicted and postdicted scores versus actual Exam 1 scores by subgroups.

(Adapted from Hacker et al., 2000.)

Perhaps even more relevant to education is how well students predict their performance in the classroom. Again, the general finding is overconfidence (for a review, see Hacker, Bol, & Keener, 2008), although a study by Hacker, Bol, Horgan, and Rakow (2000) qualifies this general conclusion in an important way by suggesting that objective achievement plays an important role in predictive accuracy. Participants in this study were undergraduate students in an educational psychology class who were asked to predict their test performance before they began the test. To examine predictive accuracy, the students were first divided into high- and low-performers based on their past test performance. As illustrated in Figure 3, high-performers exhibited fairly accurate predictions and were slightly *under*confident; whereas, low-performers showed low accuracy and were highly *over*confident, a pattern that has since been replicated (see, e.g., Grimes, 2002; Miller & Geraci, 2011). The particularly troubling aspect of this finding is the overconfidence of low-achievers—called the *unskilled-but-unaware effect* (Kruger & Dunning, 1999)—as it suggests that those students who could profit the most from remedial interventions may be oblivious to their need for such measures.

With the rising use of technology in the classroom, studies involving learners' perceptions of that technology are of increasing interest to metamemory researchers. At a broad level, Serra and Dunlosky (2010) found evidence for what they termed the *multimedia bias*—that is, the belief of learners that they learn better from text accompanied by images than from text without images. When those images are helpful (e.g., relevant diagrams), this belief aligns with actual learning, a finding supported by multiple studies (e.g., Mayer & Anderson, 1991, 1992). When those images are unhelpful, however (e.g., photographs of lightning strikes), then, contrary to participants' beliefs, learning is no different between the passage with images and the passage without images. Similarly, animations can also cause learners to experience a misleading sense of fluency. As with static images, there are situations in which animations improve learning (e.g., Hoffler & Leutner, 2007), but even more problematic, there are also situations in which static or dynamic images impair learning (e.g., Harp & Mayer, 1998). As is the case with the other illusions of competence we have discussed, however, even when multimedia presentations impair learning, learners believe that this type of presentation is helpful (Paik & Schraw, 2013).

That learners tend to believe that they learn better when the to-be-learned material is presented for study with images or animations is another example of an illusion almost certainly due to fluency. In a survey of students' perceptions of multimedia learning, one of the highest-rated characteristics of multimedia instructional materials was that they "allowed people to learn with no effort" (Antonietti, Colombo, & Lozotsev, 2008). We see such a response as a strong indication that fluency is at the root of the multimedia bias and likely to lead to some misguided decisions on the part of students in self-regulated learning environments. For example, when given the choice of what kind of captions, if any, to include in an animated, narrated lesson, students prefer to have the full narration reproduced in on-screen captions, even though research shows that full-text on-screen captions result in worse learning than no captions or abridged captions (Mayer, Heiser, & Lonn, 2001; Yue, Bjork, & Bjork, 2013).

Learners in the classroom, like those in the laboratory, may also be susceptible to how easily information is processed during learning when assessing their own competence. In a recent example of the power of fluency on JOLs, Carpenter, Wilford, Kornell, and Mullaney (2013) had participants watch a video of an instructor giving a scientific lecture on why most calico cats are female. In the *fluent* condition, the instructor delivered the content while standing upright, maintaining eye contact, and without using notes; in the *disfluent* condition, the instructor displayed the opposite behaviors—she was hunched over the podium, did not maintain eye contact, and read the lecture from notes. Immediately following the video, participants predicted the amount of information that they would remember on a test to be administered approximately 10 minutes later. Participants in the fluent condition predicted they would remember about twice as much material as did the participants in the disfluent condition. Actual learning, however, was not affected by the fluency manipulation. Thus, the fluency of a lecture can produce a robust illusion of competence among learners and, as the authors speculate, may also give rise to instructors misjudging the effectiveness of their own lectures. This latter issue—how well people can predict the performance of *others*—is addressed next.

Predicting Others' Performance

In a general argument for the importance of accurately judging the knowledge of others, Nickerson (1999) stated that "A speaker who overestimates what his or her listeners know may talk over their heads; one who underestimates their knowledge may, in the interest of being clear, be perceived as talking down to them. Both types of misjudgment work against effective and efficient communication" (p. 737). This notion is especially relevant in educational contexts in which instructors need to accurately gauge the aptitude of their pupils in order to foster effective learning in the classroom. For example, instructors must make decisions about how or the level at which certain concepts should be explained, the amount of time that should be devoted to their explanation, and when to stop for questions. Such choices are informed, to some extent, by instructors' awareness of their students' understanding of the material before, during, and after instruction.

In general, people tend to use their own subjective experiences as a proxy to predict others' performance, assuming that their own experiences are indicative of that of others (for a review, see Nickerson, 1999). This idea is exemplified in the now-classic dissertation work of Newton (1990). In her study, participants were assigned to be either "tappers" or "listeners." Tappers were asked to tap out the rhythm of popular songs (e.g., "Happy Birthday to You") on the table, after which time they were asked to assess the likelihood that the listeners would correctly identify the song, which the listeners attempted to do. Tappers predicted that listeners would correctly identify the tapped jingles 50% of the time. How often did listeners actually identify the correct tune? Less than 3% of the time! The 47% disparity between predicted and actual correct identification can be attributed to the so-called 'curse of knowledge' (see Camerer, Lowenstein, & Weber, 1989), which refers to the phenomenon that people are sometimes unable to ignore privately held information when making predictions about others. Stated differently, once we are informed, it is difficult for us to imagine what it is like to be *un*informed. In Newton's study, the tappers seemed to have had difficulty discounting inside information (the song titles) when predicting the listeners' performance, and thus they vastly overestimated the success rate of the listeners.

That we use our own subjective experience as a basis for predicting others' performance has also been demonstrated in the realm of problem solving. Kelley and Jacoby (1996), for example, asked participants to predict the likelihood that others would be able to solve anagrams. When participants were required to solve the anagrams themselves before making a prediction for others (e.g., *FSCAR -???*), participants used their response latencies (i.e., how long it took them to solve the anagrams) as a basis for predicting how difficult it would be for others to solve them. Under these conditions, a high correlation was found between participants' judged difficulty of the anagrams and their actual difficulty (established by pre-experimental norms); thus, using subjective difficulty in this context was a valid predictor of others' performance. When, however, the solutions were presented alongside the anagrams (e.g., *FSCAR—SCARF*), the correlation between judged and actual difficulty was substantially reduced, suggesting that denying participants the subjective experience of solving the anagrams forced them to use other, less diagnostic information when judging their difficulty. Again, it appears that people use their own experiences when predicting the performance of others, a tendency that may produce accurate predictions in some cases, but not in others.

How well do people predict others' memory performance? Surprisingly, this question, while important both theoretically and practically, has drawn scant empirical attention. In the first investigation of making JOLs for others, Vesonder and Voss (1985) found that individuals given access to the same information as the learner can predict the learner's performance as accurately as the learner. Similarly, Koriat et al. (2004) showed that people made similar predictions for others as they did for themselves when anticipating the effects of retention interval on memory performance, yet another finding that accords with the notion that people use their own experiences as a guide to predict the experiences of others. In one of the only other studies concerned with making JOLs for others, however, Koriat and Ackerman (2010) showed that people do not always make decisions in this way. In their paradigm, participants in a *Self* condition studied paired-associates at their own pace, making standard item-by-item JOLs after each pair. In a separate condition—the *Other* condition—participants watched a video of another person studying each pair for varying amounts of time before making JOLs regarding the likelihood that the person in the video would remember each item. Consistent with previous work on monitoring one's own learning during self-paced study (e.g., Koriat, Ma'ayan, & Nussinson, 2006), participants in the *Self* condition demonstrated an inverse relationship between study time and JOLs—that is, their JOLs decreased as the time they took to study an item increased. In the *Other* condition, however, the opposite pattern was found: Participants predicted that the longer someone else spent studying an item, the better their memory for that item would be. On the basis of this pattern, the authors argued that when people are making predictions for themselves during self-paced study, they use their own study time as an indicator of item difficulty, and consequently they give lower JOLs for items to which they devote relatively more time. When people make memory predictions for others in this situation, however, they adopt a memorizing-effort heuristic, in which later memory performance is assumed to be positively related to the amount of time spent studying. Clearly, more research is needed to identify how, and to what degree of accuracy, people forecast the memorial experiences of others, an issue that is particularly relevant in educational settings where instructors must make myriad decisions about what is best for their students' learning.

Quantifying Monitoring Accuracy: Calibration versus Resolution

Thus far, we have discussed research showing that monitoring accuracy—whether predicting one's own memory or that of others—is variable and can depend on numerous factors. Before proceeding further with our consideration of research concerned with monitoring accuracy, however, it is important to point out that such accuracy can be measured in various ways. Conceptually, of course, accuracy in the current context refers to the degree to which subjective judgments match with actual performance, but this relationship can be calculated in two different, but equally important, ways. First, *calibration* (also termed *absolute accuracy*) refers to the overall correspondence between the average prediction value and the average value of actual performance. Based on the discrepancy between these values, over- or underconfidence can be determined. To illustrate, a student might anticipate achieving an overall score of 90% on an exam but actually earn only a score of 70%, demonstrating overconfidence. *Resolution* (also termed *relative accuracy*), on the other hand, refers to the extent to which material given high predictions are associated with high performance and material given low predictions are associated with low performance, and it is most commonly measured via within-person gamma correlations (Nelson, 1984; but see Benjamin & Diaz, 2008; Masson & Rotello, 2009, for alternatives). For example, a given student might believe that material from one chapter has been learned better than material from another chapter. If correct (i.e., the student performs better on the former than on the latter material), the student's resolution is high.

Stressing the differences between calibration and resolution is not a trivial endeavor (for further discussion, see Thiede, Mueller, & Dunlosky, this volume), as these measures of accuracy are dissociable—that is, in the same experimental paradigm, calibration might be high, while resolution might be low, or vice versa. For example, Koriat, Sheffer, and Ma'ayan (2002) showed that, with practice over multiple study-test trials, calibration worsens whereas resolution improves. With respect to calibration, learners showed overconfidence in their JOLs for paired associates on the first trial but underconfidence during subsequent trials, a general finding termed the *underconfidence-with-practice (UWP) effect*. In terms of resolution, however, the learners' JOLs became more sensitive at discriminating between items they did and did not remember. This example illustrates the important point that calibration and resolution are not the same and that any conclusions regarding metamemorial accuracy will necessarily depend on how accuracy is defined. It is our view that both types of accuracy should be evaluated whenever possible, but that the specific research question at hand will ultimately determine which measure of accuracy will provide the most purchase in a given situation.

Controlling One's Own Learning

Accurately monitoring one's own learning, or the learning of others, is crucial because people act on such subjective assessments (see Nelson & Narens, 1990). Learners monitor, or reflect upon, their own learning and then use this monitoring to control, or regulate, their subsequent behavior. For example, a student, feeling unsure about the effectiveness of a study session, might choose to study the material a second time or seek the help of others. Likewise, an instructor who senses confusion among his or her students might devote extra time to explain the misunderstood concept or attempt to explain it differently. The general assumption that subjective experience plays a causal role in determining behavior was met with a fair amount of criticism in the past (e.g., Nisbett & Wilson, 1977; Reder, 1987), but it is our impression that this assumption is championed by most metamemory researchers in the field today. In this section, we survey the current landscape with respect to the notion that monitoring affects control and summarize the leading theoretical candidates for explaining this relationship, underscoring the educational implications along the way (for further details on theories of control, see Kornell, this volume).

We begin this section by briefly describing work showing that simply having control over one's own learning can benefit learning. As an example of such research, Kornell and Metcalfe (2006) investigated whether learners, when permitted to manage their own learning, profit from having such control (see also Nelson, Dunlosky, Graf, & Narens, 1994). After studying to-be-remembered material—in this case, word pairs and answers to general knowledge questions—participants were asked to decide whether they would like to restudy the material if given the opportunity. Crucially, in some conditions, restudy choices were honored (i.e., the items that were chosen for restudy were, in fact, presented for restudy); whereas, restudy decisions were not honored in other conditions, either by having participants restudy the items that were *not* chosen to be restudied or by randomly picking the items to be restudied. Across three experiments, later learning was enhanced when restudy choices were honored as compared to when they were dishonored. Thus, it is clear that metamemorial processes impact learning rather than acting as mere epiphenomena.

A wealth of metamemory research has focused more sharply on how learners use their subjective evaluations of their own learning as the basis for subsequent study decisions (for a recent review, see Dunlosky & Ariel, 2011), but surprising, it was not until relatively recently that metamemory researchers properly tested the assumption that monitoring plays a *causal* role in controlling behavior. Given the importance of this work, we describe two relevant studies in turn. First, Metcalfe and Finn (2008) had participants study word pairs either once on Trial 1 and three times on Trial 2 (the 1-3 condition) or three times on Trial 1 and once on Trial 2 (the 3-1 condition). After the last presentation of each pair on Trial 2, JOLs were elicited. Across two experiments, it was shown that (a) the 1-3 participants gave lower JOLs than did the 3-1 participants and (b) that the 1-3 participants also chose to restudy more of the pairs when given the opportunity than did the 3-1 participants. Critically, this pattern held despite equivalent levels of final recall between the conditions. That is, regardless of the fact that in each condition the pairs were studied the same number of times (4), JOLs and study behavior differed between conditions. Furthermore, as a subsequent experiment showed, this finding could not be attributed to Trial 1 recall performance. Thus, these findings provide strong support for the idea that JOLs are directly related to study behavior.

Second, Rhodes and Castel (2009) have reported similar findings (see also Soderstrom & Rhodes, 2014). Participants listened to a list of single words presented in a loud or quiet volume, making JOLs and restudy decisions after each word. A free recall test was then administered. Participants demonstrated an illusion of competence such that higher JOLs were given for loud words relative to quite words despite equivalent recall performance as a function of volume. Crucially—and aligning with the JOL reports—quiet words were also chosen for restudy more often than loud words. Thus, consistent with the results reported by Metcalfe and Finn (2008), the Rhodes and Castel study demonstrated that monitoring (JOLs) directly influences control processes (restudy choices) independently of objective recall performance. That JOLs impacted study behavior despite equivalent levels of later recall is critical for establishing a causal link between monitoring and control. Had performance differed, it could have been argued that the items that were chosen more often for restudy were chosen on an objective basis—namely, that their memory traces were weaker. Instead, that different restudy patterns were chosen for items having equivalent memory traces demonstrates that it is the learners' subjective experiences that are driving their study behavior choices.

In general, then, it is now clear that a direct relationship exists between monitoring and control processes, but more specific theories regarding this relationship are needed if we are to predict how learners use their subjective experiences to control their behavior. Three dominant theories, which we now discuss in the temporal order in which they were put forth, attempt to capture the monitoring-control relationship. First, the *discrepancy-reduction model* (Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999) contends that learners will study difficult-to-learn material (or at least material perceived to be difficult) longer than easy material in an attempt to reduce the discrepancy between what has been learned and what is sought to be learned. As Dunlosky and Thiede (1998) stated, "An item will be continued to be studied… until the person's perceived degree of learning meets or exceeds the norm of study" (p. 1024). This model thus predicts a negative correlation between study time and judged difficulty, a relationship that was indeed found in most of the early examinations of study-time allocation (for a review, see Son & Metcalfe, 2000). The notion that students spend more of their time studying relatively difficult material seems judicious and may lead to positive learning outcomes, but such a strategy may also be, in some cases, counterproductive. If the material is exceedingly difficult to learn, for example, the learner may demonstrate a *labor-in-vain effect* (Nelson & Leonesio, 1988), studying the material for exorbitant amounts of time with little or no payoff in actual learning. This misguided investment in study time would also divert valuable time away from material that *is* attainable and could thus lead to suboptimal learning outcomes, especially when overall study time is constrained (e.g., when a student schedules hour-long study sessions).

To accommodate later research showing that learners' study decisions can hinge on whether time constraints are imposed (e.g., Son & Metcalfe, 2000), a second theory attempting to capture the monitoring-control relationship—the *region of proximal learning model* (Metcalfe & Kornell, 2005)—was developed. This model asserts that learners study material until the benefits of studying can no longer be perceived, or until the rate of learning is substantially reduced. Indeed, Metcalfe and Kornell (2005) provided compelling evidence that it is not the most difficult material that is studied the longest (as predicted by the discrepancy reduction model), but rather the *moderately* difficulty material. Moderately difficult material, it was argued, is associated with the longest duration of perceived learning benefits (i.e., rate of learning) compared to easy and extremely difficult material. Easy material is learned quickly, and thus participants are led to terminate their study of such material quickly; extremely difficult material is too difficult to learn, which also leads to relatively quick study time termination. In both cases—for easy and extremely difficult material—the perceived rate of learning is quickly reduced, thus leading to relatively less study time compared to moderately difficult material. In short, the region of proximal learning model suggests that learners manage their study efforts by determining whether progress is being made during learning.

Finally, the *agenda-based regulation model* (Ariel, Dunlosky, & Bailey, 2009) suggests that task goals can also impact study decisions. Ariel et al. (2009) showed that when it came to restudy choices, importance (operationalized as arbitrary point values) superseded item difficulty as the basis for these decisions, leading to the conclusion that "learners assess task constraints prior to study and then construct an agenda that aims to efficiently achieve the current task goals within those constraints" (p. 38). Given that learners outside of the laboratory certainly have learning goals and, undoubtedly, prioritize their study schedules according to those goals, this account is particularly relevant within educational contexts. It is our view, however, that the most comprehensive theory regarding the relationship between monitoring and control will incorporate the ideas from all three accounts.

Optimizing Self-Regulated Learning

Self-regulated learning—that is, learning that is initiated and managed by the learner—is becoming an increasingly important topic of interest in the laboratory and within educational circles because technological advances, including the advent of online courses, are producing an ever-growing number of situations in which learning is occurring outside of formal classroom or training settings. As such, it is critical that research not only illuminate when learners are misled during self-regulated learning, but also identify how such errors could be corrected in order to optimize self-regulated learning. Fortunately, various ways of enhancing self-regulated learning have been identified (for a review, see Bjork, Dunlosky, & Kornell, 2013), several of which we now summarize.

Broadly speaking, enhancing self-regulated learning requires that learners become sophisticated in terms of understanding how learning and memory operate. Students and instructors need to know, for example, which types of learning strategies foster long-term learning (e.g., testing oneself, spacing study sessions) before we can expect students to adopt, and instructors to encourage, their use (for a review of the utility of various study strategies, see Dunlosky et al., 2013). As we discussed earlier, educational interventions have been shown to increase the likelihood that students correctly identify empirically validated learning strategies (McCabe, 2011) and that high-performing students are more likely to use effective strategies compared to their lower-performing peers (Hartwig & Dunlosky, 2012). Ideally, then, students would learn about effective study strategies early in their education.

That people are not, without targeted instruction, knowledgeable of what effective learning entails and, in fact, hold beliefs about how best to learn that run counter to empirical evidence implies a counterintuitive nature of learning and memory principles. How is it that, given our vast life experiences in which we have had to learn new knowledge and new skills in a vast variety of situations, we do not come to have an accurate understanding of how something as fundamental as learning works? One answer provided by Soderstrom and Bjork (2013, in press; see also Bjork, 1999) is that people have a tendency to conflate short-term *performance* with long-term *learning* when, in fact, there is overwhelming evidence that learning and performance are dissociable. Paradoxically, conditions that impair short-term performance are often beneficial on the long-term—Bjork (1994) termed these conditions *desirable difficulties*—and conversely, conditions that produce exceptional short-term performance often lead to impaired long-term learning. Revisiting a previous example, restudying information often trumps testing (i.e., retrieval practice) on immediate memory tests, but this pattern typically reverses on delayed tests, such that testing enhances long-term retention relative to restudying. More germane to our point is that people *think* that restudying is more beneficial to learning than testing, perhaps because restudying produces a subjective sense of fluency and rapid improvements during encoding or acquisition. Thus, optimizing self-regulated learning will also require that learners become sensitive to the general notion that short-term performance is not a reliable index of long-term learning and that difficult, or challenging, learning conditions often lead to enhanced long-term learning. As recently advocated by Bjork and Bjork (2011), students need to learn how to make things hard on themselves, but in a good way, during acquisition or study—that is, how to create the type of desirable difficulties during their study that will enhance their long-term learning and transfer.

Predictive judgments, like JOLs, also seem to be largely based on acquisition factors, such as how easily information is processed during encoding, the effects of which can quickly dissipate and be negligible for actual learning. Given that such assessments impact subsequent behavior, however, self-regulated learning can also profit from methods that improve learners' accuracy in predicting their future performance, a couple of which we have already discussed—namely, delaying judgments and generating keywords before assessing comprehension. Another way of improving the veracity of prospective judgments is to encourage learners to think about retrieval operations when making them. For example, McCabe and Soderstrom (2011) developed and solicited judgments that encouraged people to think about *how* information will be remembered later by asking participants to predict whether recollective details associated with an event would be recalled. Across a range of experiments using different instructions, materials, and outcome measures, it was found that recollection-based prospective judgments were more accurate than confidence-based JOLs. In a follow-up study, Soderstrom and Rhodes (2014) showed that such recollection-based metamemorial judgments are less susceptible to the influence of irrelevant, fleeting perceptual information—in this case, the auditory intensity in which to-be-remembered information is presented—compared to JOLs and, especially important, that this relative immunity was reflected in learners' restudy decisions. That is, encouraging learners to base their prospective judgments on retrieval factors improved monitoring accuracy, which, in turn, led to more informed study decisions.

Finally, optimizing self-regulated learning can, in part, be achieved by equipping learners with a testing experience. DeWinstanley and Bjork (2004; see also Bjork & Storm, 2011), for example, had participants first read a passage that included both to-be-generated and to-be-read information, which resulted in learners experiencing the generation effect—namely, that generated material was better remembered on the criterion test than read material. Next, participants read a new passage—again containing to-be-generated and to-be-read information—but this time no generation effect materialized because learners applied a generation-based strategy to the to-be-read information. Critically, participants who were denied the experience of the generation effect did not show the enhanced encoding of subsequent to-be-read items. More recently, Soderstrom and Bjork (2014) showed that interim tests inform learners of gaps in their knowledge and, consequently, enhance the efficiency and effectiveness of subsequent study behavior compared to when no such interim test is taken. Moreover, this test-enhanced self-regulated learning transferred to nontested material when restudied in the context of material that *was* tested. To illustrate the educational implications of this work, consider an instructor who gives a quiz at the end of a lecture covering half of the lecture's content. These data suggest that when students restudy that lecture content on their own, their self-regulated learning will be enhanced—for both the quizzed and nonquizzed material—compared to when no quiz was given. Consistent with this idea, Bjork, Little, and Storm (2014) found that students in a large undergraduate course in which quizzes were given following lectures not only performed better on the course final exam on those questions but also on related questions—that is, questions that had not themselves appeared on the previous quizzes but that asked about information related to the material quizzed earlier. Other research showing that the benefits of testing some information can extend to learners' ability to answer questions about related information on a delayed test have recently been reported on the basis of laboratory-based studies by Butler (2010) and classroom-based studies by McDaniel, Thomas, Agarwal, McDermott, and Roediger (2013).

Conclusions and Future Directions

In this chapter, we have presented a selective review of the current literature on metamemory research with regard to its educational implications. A wide range of empirical methods—from the laboratory to the classroom—has revealed important insights into what students believe about learning and how learners assess and regulate their ongoing learning. Broadly speaking, people hold faulty beliefs about how learning and memory operate, often endorsing and employing study strategies that lead to relatively poor long-term retention. Likewise, learners' ongoing assessments of their own learning are frequently inaccurate, producing illusions of competence that are rooted, in part, in people's tendency to use current performance as an index of long-term learning. Faulty beliefs and inaccurate monitoring can, in turn, lead to suboptimal self-regulated learning. Fortunately, several ways of enhancing self-regulated learning have been identified, including targeted instruction on learning and memory topics, and equipping learners with a testing experience to expose gaps in their knowledge. Given that learning is increasingly occurring in unsupervised settings, we believe that even more empirical attention should be devoted to uncovering ways of optimizing self-regulated learning.

In educational contexts, of course, it is not only important that students accurately monitor their own learning, but also that instructors accurately gauge the aptitude of their students so as to appropriately tailor their instruction. Unfortunately (and surprisingly), there is a paucity of research regarding the extent to which people can accurately judge the learning of others, the cues that are used to form such judgments, and the consequences of those subjective assessments. Thus, we encourage researchers to investigate these important issues. We briefly summarized work in other domains showing that people are often unable to discount privately held information when making behavioral forecasts of others—the signature of the so-called *curse of knowledge*. If such a 'curse' is also shown to contaminate metamemory judgments about others, then future research should also be aimed at discovering ways to ameliorate this bias.

Finally, the late Ulrich Neisser (1976) stated that one of our major goals, as researchers, should be “to understand cognition in the context of natural purposeful activity” (p. 4). We agree with this statement and thus also encourage metamemory researchers to continue to examine metamemorial processes in the classroom and with educationally relevant materials. Such research, while certainly challenging, is crucial for us to pursue if we are to achieve the goal of having our research findings translated into action by educators and students alike.

Author Note

Nicholas C. Soderstrom, Carole L. Yue, and Elizabeth Ligon Bjork, Department of Psychology, University of California, Los Angeles.

Grant 29192G from the James S. McDonnell Foundation supported the writing of this chapter.

Correspondence concerning this article should be addressed to Nicholas C. Soderstrom, Department of Psychology, 1285 Franz Hall, UCLA, Los Angeles, CA 90095-1563. E-mail: nsoderstrom@psych.ucla.edu.

References

Antonietti, A., Colombo, B., & Lozotsev, Y. (2008). Undergraduates’ metacognitive knowledge about the psychological effects of different kinds of computer-supported instructional tools. *Computers in Human Behavior*, 24, 2172–2198.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*, 126-131.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, 138, 432–447.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28, 610–632.

Bell, K. E., & Limber, J. E. (2010). Reading skill, textbook marking, and course performance. *Literacy Research and Instruction*, 49, 56–67.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, 31, 297–305.

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.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York, NY: Psychology Press.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201–210.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, *3*, 165-170.

Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1113–1124.

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

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds.), Attention and performance XVII. *Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118–1133.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, 97, 1232–1254.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*, 1350–1356.

Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills*, 96, 54–56.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, 36, 429–437.

Castel, A. D., McCabe, D. P., & Roediger, H. L., III. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, 14, 107–111.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380.

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

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

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2010). Fortune favors the **bold** (*and the italicized*): Effects of disfluency on educational outcomes. *Cognition*, 118, 111–115.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed.), *Psychology of learning and motivation*, Vol. 54 (pp. 103–140). San Diego, CA: Elsevier Academic Press.

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

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.

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.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). New York: Dover Publications. (Original work published 1885).

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.

Fowler R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied*

Psychology, 59, 358–364.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 702–718.

Grimes, P. W. (2002). The overconfidence principles of economics students. An examination of a metacognitive skill. *The Journal of Economic Education*, 33, 15–30.

Hacker, D. J., Bol, L., Horgan, D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92, 160–170.

Hacker, D. J., Bol, L., & Kenner, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlsoky, & R. A Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429–455). New York, NY: Psychology Press.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414–434.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of verbal learning and verbal behavior*, *6*, 685–691.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19, 126–134.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 22–34.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722–738.

Jang, Y., & Nelson, T. O. (2005). How many dimensions underlie judgments of learning and recall? Evidence from state-trace methodology. *Journal of Experimental Psychology: General*, 134, 308–326.

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

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, 35, 157–175.

Koch, A. (2001). Training metacognition and comprehension of physics texts. *Science Education*, 85, 758–768.

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., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and Cognition*, 19, 251–264.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 187–194.

Koriat, A., Bjork, R. A., Sheffer, L., & Barr, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643–656.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationship between monitoring and control in metacognition: Lessons for the cause-and effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135, 36–69.

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. (2011). Failing to predict future changes in memory: A stability bias yields long-term overconfidence. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festchrift in honor of Robert A. Bjork* (pp. 365–386). New York, NY: Psychology Press.

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

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, 19, 585–592.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating forgetting. *Journal of Experimental Psychology: General*, 138, 449–468.

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.

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

Mackenzie, A. M. (1994). Examination preparation, anxiety and examination performance in a group of adult learners. *International Journal of Lifelong Education*, 13, 373–388.

Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Findings and implications for education. In T. Perfect and B. Schwartz (Eds.), *Applied metacognition* (pp. 39–67). Cambridge, England: Cambridge University Press.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97, 723–731.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 509–527.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83, 484–490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444–452.

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

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

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93, 187–198.

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

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

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39, 462–476.

McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKs). *Journal of Experimental Psychology: General*, 140, 605–621.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L., III. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*, 360-372.

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., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware. Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 502–506.

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people’s beliefs about memory? *Journal of Memory and Language*, *70*, 1-12.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: Freeman.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133.

Nelson, T. O., & 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., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207–213.

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. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York, NY: Academic Press.

Newton, L. (1990). *Overconfidence in the communication of intent: Heard and unheard melodies*. Unpublished doctoral dissertation. Stanford University, Stanford, CA.

Nickerson, R. (1999). How we know—and sometimes misjudge—what others know: Imputing one’s own knowledge to other. *Psychological Bulletin*, 125, 737–759.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.

Paik, E. S., & Schraw, G. (2013). Learning with animations and illusions of understanding. *Journal of Educational Psychology*, 105, 278–289.

Peterson, S. E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction*, 31, 49–56.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, 19, 90–138.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16, 550–554.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning (JOLs) on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, 137, 131–148.

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

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

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief that diagrams improve learning from text. *Memory*, 18, 698–711.

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition*, *1*, 145–151.

Soderstrom, N. C., & Bjork, R. A. (2013). Learning versus performance. In D. S. Dunn (Ed.), *Oxford bibliographies online: Psychology*. New York, NY: Oxford University Press.

Soderstrom, N. C., & Bjork, R. A. (in press). Learning versus performance: An integrative review. *Perspectives on Psychological Science*.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time, *Journal of Memory and Language*, *73*, 99-115.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1236–1242.

Soderstrom, N. C., & Rhodes, M. G. (2014). Metacognitive illusions can be reduced by monitoring recollection during study. *Journal of Cognitive Psychology*, 26, 118–126.

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

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, 5, 315–316.

Stordahl, K. E., & Christensen, C. M. (1956). The effect of study techniques on comprehension and retention. *Journal of Educational Research*, 49, 561–570.

Sungkhasette, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, 18, 973–978.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, 20, 356–363.

Thiede, K., & Anderson, M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.

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.

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-learning-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1264–1269.

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language*, 24, 363–376.

Wade, S. E., & Trathen, W. (1989). Effect of self-selected study methods on learning. *Journal of Educational Psychology*, 81, 40–47.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, 105, 266–277.

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, 41, 229–241.

Yue, C. L., Storm, B. C., Kornell, N., & Bjork. E. L. (in press). Highlighting and its relation to distributed study and students' metacognitive beliefs. *Educational Psychology Review*.

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

Nicholas C. Soderstrom

Department of Psychology, University of California, Los Angeles

Carole L. Yue

Department of Psychology, University of California, Los Angeles

Elizabeth Ligon Bjork

Department of Psychology, University of California, Los Angeles