Learning Versus Performance: An Integrative Review

Article *in* Perspectives on Psychological Science · March 2015

DOI: 10.1177/1745691615569000 · Source: PubMed

Robert A. Bjork
University of California, Los Angeles

Perspectives on Psychological Science
2015, Vol. 10(2) 176–199
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691615569000
pps.sagepub.com

Learning Versus Performance: An Integrative Review

Nicholas C. Soderstrom and Robert A. Bjork
Department of Psychology, University of California, Los Angeles

Abstract

The primary goal of instruction should be to facilitate long-term *learning*—that is, to create relatively permanent changes in comprehension, understanding, and skills of the types that will support long-term retention and transfer. During the instruction or training process, however, what we can observe and measure is *performance*, which is often an unreliable index of whether the relatively long-term changes that constitute learning have taken place. The time-honored distinction between learning and performance dates back decades, spurred by early animal and motor-skills research that revealed that learning can occur even when no discernible changes in performance are observed. More recently, the converse has also been shown—specifically, that improvements in performance can fail to yield significant learning—and, in fact, that certain manipulations can have opposite effects on learning and performance. We review the extant literature in the motor- and verbal-learning domains that necessitates the distinction between learning and performance. In addition, we examine research in metacognition that suggests that people often mistakenly interpret their performance during acquisition as a reliable guide to long-term learning. These and other considerations suggest that the learning–performance distinction is critical and has vast practical and theoretical implications.

Keywords

learning, performance, memory, instruction, training, motor learning, verbal learning

Whether in the classroom or on the field, the major goal of instruction is, or at least should be, to equip learners with knowledge or skills that are both durable and flexible. We want knowledge and skills to be durable in the sense of remaining accessible across periods of disuse and to be flexible in the sense of being accessible in the various contexts in which they are relevant, not simply in contexts that match those experienced during instruction. In other words, instruction should endeavor to facilitate *learning*, which refers to the relatively permanent changes in behavior or knowledge that support long-term retention and transfer. Paradoxically, however, such learning needs to be distinguished from *performance*, which refers to the temporary fluctuations in behavior or knowledge that can be observed and measured during or immediately after the acquisition process.

The distinction between learning and performance is crucial because there now exists overwhelming empirical evidence showing that considerable learning can occur in the absence of any performance gains and, conversely, that substantial changes in performance often fail to translate into corresponding changes in learning. Perhaps even more compelling, certain experimental manipulations have been shown to confer opposite effects on learning and performance, such that the conditions that produce the most errors during acquisition are often the very conditions that produce the most learning. Such results are regularly met with incredulity, whether in the context of metacognitive research in which people are asked to make judgments about their own learning or during informal conversations with researchers, educators, and students. It is, however, the counterintuitive nature of the learning–performance distinction that makes it so interesting and important from both practical and theoretical perspectives.

We provide the first integrative review of the evidence that bears on the critical distinction between learning, as measured by long-term retention or transfer, and performance, as measured during acquisition. We attempt to

---

Corresponding Author:
Nicholas C. Soderstrom, Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA 90095-1563
E-mail: nsoderstrom@psych.ucla.edu

synthesize research from both the motor- and verbal-learning domains, as well as relevant work in metacognition. We note, however, that a number of other articles provide an introduction to the learning versus performance distinction and summarize key findings that illustrate the distinction (e.g., R. A. Bjork, 1999; Christina & Bjork, 1991; Jacoby, Bjork, & Kelley, 1994; Kantak & Winstein, 2012; Lee, 2012; Lee & Genovese, 1988; Schmidt & Bjork, 1992; Schmidt & Lee, 2011; Wulf & Shea, 2002). As well, we (Soderstrom & Bjork, 2013) have published an annotated bibliography that is slated to be updated annually so as to keep researchers, educators, and others abreast of the newest research relevant to the topic.

Our review begins by presenting the foundational research on which the learning–performance distinction rests—specifically, the early work on latent learning, overlearning, and fatigue—and we then highlight the corresponding conceptual distinctions made by learning theorists at that time. Next, we discuss various experimental manipulations from both the motor- and verbal-learning domains that have resulted in dissociations between learning and performance. We then summarize research findings in the domain of metacognition that demonstrate that learners are prone to interpreting performance during acquisition as a valid index of learning, which can lead not only to misassessments of the degree to which learning has happened but also to learners preferring poorer conditions of learning over better conditions of learning. Finally, we present several current theoretical perspectives that can accommodate the difference between learning and performance.

Foundational Studies

Studies conducted decades ago necessitated the distinction between learning and performance by showing that considerable learning could occur in the absence of changes in performance. For example, rats' learning of a maze could be enhanced by permitting a period of free exploration in which their behavior seemed aimless (i.e., performance was irregular); additional practice trials provided after performance was at asymptote ("overlearning") resulted in slowed forgetting and more rapid relearning; and when fatigue stalled performance of to-be-learned motor tasks, learning could still transpire. This section reviews these foundational studies.

Latent learning

Latent learning is defined as learning that occurs in the absence of any obvious reinforcement or noticeable behavioral changes. Learning is said to be "latent," or hidden, because it is not exhibited unless a reinforcement of some kind is introduced to reveal it. Consider, for example, a person who recently moved to a new city and, apprehensive about driving, decides to ride the city bus each day to work. Riding the bus day after day, the route would be learned through observation, but such learning would only be evident if an incentive was present that required it—say, when it was necessary for the person to drive to work on his or her own. The early findings of latent learning were intriguing and controversial because they challenged the widely held assumption that learning could occur only in the presence of reinforcement. For a classic review of the early latent learning studies, we recommend Tolman (1948), in which the concept of "cognitive maps" was introduced, a term that refers to the mental representation of one's spatial environment.

Although first demonstrated by Blodgett (1929), Tolman and Honzik (1930) are credited for providing what is now considered the classic experiment on latent learning, the results of which are reported in most textbooks on learning and memory. In their experiment, which is essentially a replication of Blodgett's, three groups of rats were placed in a complex T-maze every day for a total of 17 days. One group of rats was never reinforced for reaching the goal box—they were simply taken out of the maze when they found it—whereas another group was reinforced with food every time the goal box was reached. A third group was not rewarded for reaching the goal box until Day 11, after which time they were regularly rewarded. The results of this experiment are presented in Figure 1. Unsurprisingly, the group that made the fewest errors in finding the goal box over the 17-day period was the regularly reinforced group, and the group that was never reinforced made the most errors. Consistent with the notion of latent learning, the delayed-reinforcement group showed the same number of errors as the never-reinforced group until Day 11—the day the food was introduced—when an immediate improvement occurred, dropping their error rate to a level comparable to that of the regularly reinforced group. Thus, delaying reinforcement revealed that the rats did, indeed, learn the maze while no reinforcement was provided and their behavior seemed rather aimless. In other words, learning occurred when performance was stagnant.

The studies by Blodgett (1929) and Tolman and Honzik (1930) spurred numerous follow-up experiments on latent learning in rats, further refining our understanding of this phenomenon (see Buxton, 1940; Spence & Lippitt, 1946). Seward (1949), for example, showed that latent learning could occur after just 30 min of free exploration and, furthermore, that the amount of time spent in the maze with no reinforcement—and thus during a time when no changes in performance were discernible—was positively related to learning the maze (see also Bendig, 1952; Reynolds, 1945).

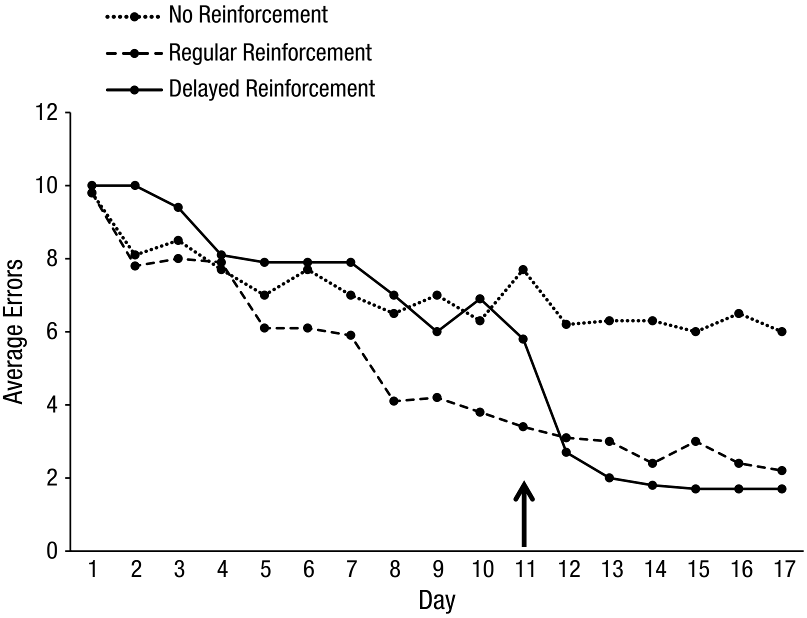

Fig. 1. Average number of errors rats committed while trying to find the goal box as a function of time and reinforcement group. The arrow above Day 11 denotes when reinforcement (food) was introduced to the delayed-reinforcement group. (Note that lower scores represent better performance and learning.) Data are adapted and approximated from Tolman and Honzik (1930).

It was also made clear decades ago that latent learning is not limited to rats. In their influential studies, Postman and Tuma (1954) and Stevenson (1954) showed that latent learning is also empirically demonstrable in humans. In Stevenson's experiment, children—some as young as 3 years old—explored a series of objects to find a key that would open a box. Critically, the explored environment also contained nonkey objects, or those that were irrelevant to the task. The question was whether the children would learn the locations of these peripheral objects during the exploration of the key-relevant objects—that is, whether the children would show latent learning. Indeed, when the children were asked to find the irrelevant, nonkey objects, they were relatively faster in doing so when those objects had been contained in the explored environment. Stevenson also found that the amount of latent learning observed in the children increased with age.

Overlearning and fatigue

Consider a violinist who continues to practice a musical piece despite already being able to perform it—that is, after acquisition performance is already at asymptote. Such continued practice on a task after some criterion of mastery on that task has been achieved is referred to as "overlearning" and can be expressed by the number of postmastery trials divided by the number of trials needed to reach mastery. For example, if the violinist practiced a piece 5 additional times after needing 10 practice trials to master it, then the degree of overlearning would be 50%. Many early studies of overlearning—starting with Ebbinghaus's (1885/1964) famous study using nonsense syllables—demonstrated the power of overlearning as a method for enhancing the long-term learning of information and skills. Referencing these findings, Fitts (1965) stated, "The importance of continuing practice beyond the point in time where some . . . criterion is reached cannot be overemphasized" (p. 195).

Krueger (1929) carried out the most frequently cited study on overlearning. In his seminal experiment, two groups of participants repeatedly studied lists of words until all of the words could be recalled. At that point, the control group was finished with the study phase, whereas participants in the overlearning group continued studying the material—in fact, they overlearned the material by 100%, meaning that they were exposed to twice as many study trials as the control group. On a retention test administered up to 28 days later, the participants in the overlearning group recalled more items than participants in the control group, who had mastered the material during the study phase but had not overlearned it. Additionally, retention increased as a function of the degree of overlearning. Subsequent research showed that overlearning aids in the retention of more complex verbal materials, such as prose passages, and accelerates the rate of relearning—that is, the amount of time required to learn the material again after some delay (e.g., Gilbert, 1957; Postman, 1962; see also Ebbinghaus, 1885/1964).

Overlearning also benefits the learning of motor skills. The year after Krueger (1929) demonstrated overlearning for words, he (Krueger, 1930) showed similar benefits for a maze-tracing task. Participants first performed the task until they reached 100% accuracy, after which they overlearned it by 50%, 100%, or 200%. As with the verbal materials, the amount of overlearning was positively related to long-term retention. Later work replicated the benefits of overlearning for simple and more complex motor skills (e.g., Chasey & Knowles, 1973; Melnick, 1971; Melnick, Lersten, & Lockhart, 1972), including the assembly and disassembly of an M60 machine gun (Schendel & Hagman, 1982). Overlearning seems to be an effective learning tool for a wide range of tasks (for a meta-analytic review, see Driskell, Willis, & Cooper, 1992).

Similar to research on overlearning, early work on fatigue suggested that learning could occur even after fatigue prevented any further gains in performance during acquisition. Adams and Reynolds (1954), for example, had basic trainees from the Air Force learn a rotary pursuit task, which requires one to manually track a target on a revolving wheel with a wand. Varying the length of rest intervals between trials showed that when fatigue limited or eliminated gains in performance, learning nonetheless occurred, as revealed by a subsequent test on the task after the fatigue had dissipated. Fifteen years later, Stelmach (1969) examined how different training schedules affect learning and performance on a ladder-climbing task. One group of participants practiced more than they rested; another group rested more than they practiced. Performance during training, which was defined as the number of rungs climbed on a given trial, favored the group that was permitted more interpolated rest. This finding is not surprising given that the other group, as a result of receiving little rest between trials, became increasingly fatigued during the training. After a delay, however, a retention test revealed that the group that received little rest caught up to the well-rested group, ostensibly demonstrating that substantial learning had occurred when fatigue had stifled any gains in short-term performance.

Corresponding conceptual distinctions

The early experiments on latent learning, overlearning, and fatigue, plus other considerations, led early learning theorists (e.g., Estes, 1955a; Guthrie, 1952; Hull, 1943; Skinner, 1938; Tolman, 1932) to distinguish between behaviors that can be observed during training, or acquisition (i.e., performance), and the relatively permanent changes that occur in the capability for exhibiting those behaviors in the future (i.e., learning). Hull used the terms *habit strength* of a response and the *momentary reaction potential* of that response; Estes, in his fluctuation model, referred to *habit strength* and *response strength*; and Skinner differentiated between *reflex reserve* and *reflex strength*. Empirically, habit strength, or reflex reserve (i.e., learning), was assumed to be indexed by resistance to extinction or forgetting, or by the rapidity of relearning, whereas momentary reaction potential, response strength, or reflex strength (i.e., performance) was assumed to be indexed by the current probability, rate, or latency of a response.

In the domain of human verbal learning, Tulving and Pearlstone's (1966) distinction between "availability" (i.e., what is stored in memory) and "accessibility" (i.e., what is retrievable at any given time) also maps, albeit not perfectly, onto learning and performance, respectively. Finally, R. A. Bjork and Bjork (1992), in an effort to account for a wide range of findings in research on human verbal and motor learning, formulated a new theory of learning in which the distinction between learning and performance is indexed by *storage strength* and *retrieval strength*, respectively. This account, as well as other contemporary theoretical perspectives regarding the learning–performance distinction, is discussed later.

Summary

The learning versus performance distinction can be traced back decades when researchers of latent learning, overlearning, and fatigue demonstrated that long-lasting learning could occur while training or acquisition performance provided no indication that learning was actually taking place. The results of latent learning studies, in particular, were both compelling and controversial at the time because they verified that, although reinforcement is necessary to reveal learning, it is not required to induce learning. In sum, this early work showed learning without performance. In the next several sections, we review more recent evidence showing that the converse is also true—specifically, that gains in performance often impede posttraining learning compared with those conditions that induce more performance errors.

Distribution of Practice

The dissociation between learning and performance has been repeatedly found by manipulating the study schedules of to-be-learned skills or information. Massing practice or study sessions—that is, practicing or studying the same thing over and over again—usually benefits short-term performance, whereas distributing practice or study—that is, separating practice or study sessions with time or other activities—usually facilitates long-term learning. This section presents, in turn, experiments from the motor- and verbal-learning domains in which the distribution of practice was shown to have differential influences on learning and performance.

Motor learning

Suppose a swimmer wishes to improve his or her front, back, and butterfly strokes. Suppose further that the swimmer's training is restricted to 1 hr per day. One training option would be to mass (or block) the different strokes by practicing each for 20 min before moving on to the next, never returning to the previously practiced strokes during that training session. Alternatively, he or she might distribute (or randomize) the practice schedule such that each stroke is practiced for 10 min before moving on to the next stroke. This schedule would permit each stroke to be revisited one more time during the training session. In this section, we review research that suggests that, whereas massing practice might promote rapid performance gains during training, distributing practice facilitates long-term retention of that skill.

Baddeley and Longman (1978) and J. B. Shea and Morgan (1979) published two classic studies that showed that distributing practice has differential effects on learning and performance of a simple motor skill. Commissioned by the British Postal Service, Baddeley and Longman investigated how to optimize postal workers' ability to type newly introduced postcodes on the keyboard. The question was whether the postal workers should learn the new system as rapidly as possible, practicing several hours per day, or whether learning would profit most if practice was more distributed. Varying the amount of practice per day and the number of days in which practice occurred, more distributed practice fostered more effective learning of the typewriter keystrokes; however, the opposite was true in regard to the efficiency in which the skill was acquired, as measured by the number of days to reach criterion versus the number of hours to reach criterion—that is, the distributed group required more days to reach any given level of performance relative to the massed group. In sum, massed practice supported quicker acquisition of the keystrokes, but distributed practice led to better long-term retention of the skill (see also Simon & Bjork, 2001).

J. B. Shea and Morgan (1979) also showed that distributing practice benefits the long-term retention of a motor skill. In their seminal experiment, participants learned three different movement patterns, each of which involved knocking over three (of six) small wooden barriers in a prescribed order. Two different practice schedules were implemented: blocked and random. In the blocked-practice condition, each of the three movement patterns was practiced for 18 trials in succession, whereas in the random-practice condition, the 18 trials of each pattern were intermingled among the trials on the other patterns in a way that was unpredictable from a participant's standpoint. Importantly, therefore, practice time for the three tasks was equated across the two different practice conditions. Of interest was how the different practice schedules affected the rapidity in which the arm movements were executed.

The results of J. B. Shea and Morgan (1979) are shown in Figure 2. First, it is clear that during acquisition, participants assigned to the blocked-practice condition performed better than those in the random-practice condition, as evidenced by shorter times required to perform the arm movements. But how well would each group retain the acquired skills? On retention tests given after 10 min and 10 days, participants were tested on each skill in either a blocked (B) or random (R) fashion, which produced four subgroups of participants: B-B, B-R, R-B, and R-R. The first letter in the pair denotes how practice was scheduled during acquisition; likewise, the second letter denotes how each group was tested. As can be seen in Figure 2, the advantage of blocked practice during acquisition was no longer evident after a delay. In fact, the pattern reversed when learning was assessed after 10 min and 10 days—that is, overall, those who initially practiced the skills in a random order exhibited the most learning. Comparing the groups that were tested in a blocked fashion (B-B vs. R-B), the study showed that random practice during acquisition was better than blocked practice during acquisition. This pattern was dramatically demonstrated when researchers compared the groups that were tested in a random order (B-R vs. R-R) on the delayed test. Thus, similar to Baddeley and Longman (1978), Shea and Morgan showed that blocking practice of several to-be-learned movement patterns facilitated acquisition performance, whereas interleaving practice of those same movements promoted long-term retention (see Lee & Magill, 1983, for a replication and extension of these findings). Although not shown in Figure 2, Shea and Morgan also found a transfer advantage of interleaving, such that participants who initially practiced the skill in a random fashion were relatively better in executing a new response pattern—that is, one that had not been practiced.

J. B. Shea and Morgan's (1979) results, together with the multiple subsequent demonstrations that interleaving separate to-be-learned tasks can enhance long-term retention, serve as one example of a broader finding, referred to by Battig (1979) as *contextual interference* effects. Primarily on the basis of findings from verbal paired-associate learning tasks (see Battig, 1962, 1972), Battig proposed that conditions during acquisition that act to increase the possible interference between

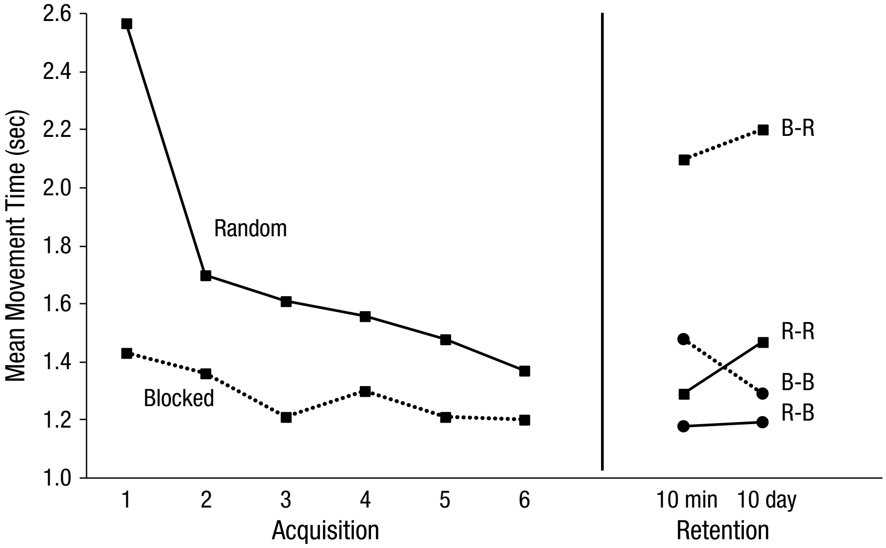

Fig. 2. Mean movement time as a function of practice schedule during acquisition and on the retention tests administered 10 min and 10 days later. For the retention data, the first letter in the pair denotes how practice was scheduled during acquisition (*B* or *R*, for *blocked* or *random*, respectively); the second letter denotes how each group was tested. (Note that lower scores represent better performance and learning.) Data are adapted and approximated from J. B. Shea and Morgan (1979).

separate to-be-learned tasks can enhance long-term retention and transfer, despite their depressing effects on performance during the acquisition process. Randomly intermixing the trials on separate to-be-learned tasks, such as, say, the forehand, backhand, and serve strokes in tennis, increases the interference between the components of those strokes but then can enhance long-term retention of those skills.

The results of J. B. Shea and Morgan (1979) spurred many follow-up studies, many of which were field based and examined more complex motor skills. In one such study, badminton players learned three different types of serves from one side of the court under blocked or randomly interleaved practice schedules. After a retention interval, the players were tested on the serves from both the same and opposite side of the court from which the serves were practiced. The blocked group performed better during training, but the interleaved group showed better long-term retention, whether tested on the same or opposite side of the court (S. Goode & Magill, 1986). Thus, not only does distributing practice enhance the retention of the specific skill that is practiced, but it also fosters better transfer of that skill—that is, the application of the skill in a different context. The learning benefits promoted by distributed practice have also been demonstrated for learning to hit pitches of different types in baseball (Hall, Domingues, & Cavazos, 1994) and piano pieces (Abushanab & Bishara, 2013) and for both children (e.g., Ste-Marie, Clark, Findlay, & Latimer, 2004) and older adults (e.g., Lin, Wu, Udompholkul, & Knowlton, 2010). For reviews of the effects of distributed practice on motor skills, both simple and complex, we recommend Lee (2012) and Merbah and Meulemans (2011).

Verbal learning

As in the motor domain, empirical evidence from verbal tasks suggests that distributing (or spacing) study opportunities benefits learning relative to massing them, a finding in the verbal literature termed the *spacing effect*. The first to demonstrate the spacing effect, Ebbinghaus (1885/1964) showed that spacing study opportunities, as opposed to massing them, rendered the material more resistant to forgetting. Decades later, now-classic articles were published on the topic (e.g., Battig, 1966; Madigan, 1969; Melton, 1970). For example, and particularly relevant to the current review, Peterson, Wampler, Kirkpatrick, and Saltzman (1963) were the first to observe that massed items are often retained better in the short term (i.e., spacing impairs *performance*), whereas spaced items are retained better over the long term (i.e., spacing enhances *learning*; see also Glenberg, 1977). Since then, hundreds of experiments have demonstrated the spacing effect to be highly robust and reliable (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1988). We now selectively review evidence of the spacing effect and how this experimental manipulation bears on the learning–performance distinction. We note that we have grouped together situations in which spacing is achieved in two different ways: (a) by inserting periods of rest or unrelated activity between repetitions of to-be-learned information or procedures; and (b) by interleaving the study or practice trials of several different—and possibly interfering—to-be-learned tasks or verbal materials. A currently active issue, however, is whether the benefits of interleaving go beyond the benefits of the spacing such interleaving introduces (see, e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012).

The majority of studies examining the spacing effect have done so using relatively simple to-be-learned materials, such as single words or paired associates. In one study, for example, high school students learned French–English vocabulary pairs (e.g., *l'avocat—lawyer*) under conditions of either massed practiced, in which the pairs were studied for 30 consecutive minutes on one day, or spaced (distributed) practice, in which the pairs were studied for 10 min on each of three consecutive days. On an initial test that was administered immediately following each practice schedule—after the 30-min study session for the massed group and after the third 10-min study session for the spaced group—virtually identical short-term performance was observed. However, on a long-term retention test administered 7 days later, participants who had spaced their study recalled more pairs than participants who had massed their study (Bloom & Shuell, 1981). Similarly, spacing study sessions, relative to massing them, can actually slow down the acquisition of foreign language vocabulary pairs but can still lead to superior retention—even over a span of several years (Bahrick, Bahrick, Bahrick, & Bahrick, 1993).

Age-related differences in the spacing effect have also been examined. For example, both younger (ages 18–25) and older (ages 61–76) adults studied unrelated paired associates (e.g., *kitten–dime*) multiple times according to either a massed or spaced presentation schedule. Using a continuous cued-recall paradigm, each item was tested after either 2 (short retention) or 20 (long retention) intervening items were presented following the item's last presentation. An unsurprising finding was that older adults performed worse, overall, compared with their younger counterparts. More interesting, and relevant to the learning–performance distinction, both age groups exhibited a spacing-by-retention-interval interaction—that is, short-term retention (i.e., performance) favored the massed items, whereas long-term retention (i.e., learning) favored the spaced items (Balota, Duchek, & Paullin, 1989).

In addition to fostering better retention of simple materials, spacing also improves the learning of more complex materials, such as prose passages (e.g., Rawson & Kintsch, 2005), and the learning of higher-level concepts, such as logic (Carlson & Yaure, 1990) and inductive reasoning (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). A particularly striking example showed that spacing various types of math problems, as opposed to massing them, facilitates learning. Participants' task was to learn how to find the geometric volume of four differently shaped objects. One group worked through the practice problems according to a blocked schedule, such that four problems for one object were attempted before moving on to four problems for the next object, and so on; the other group worked through the problems for various shapes in a randomly mixed order. Participants were then tested on the problems 1 week later. During the practice phase, participants were able to solve more of the problems if those problems were practiced in a blocked fashion—that is, massing improved performance. This pattern reversed, however, on the long-term retention test: Participants better retained the ability to solve the problems if those problems were practiced 1 week earlier in a mixed format—that is, spacing enhanced learning (Rohrer & Taylor, 2007). These results exemplify the distinction between learning and performance (see also Rohrer, Dedrick, & Burgess, 2014; Taylor & Rohrer, 2010).

Summary

Evidence from the motor- and verbal-learning domains demonstrates that long-term learning profits from distributing (spacing) the practice of to-be-learned skills or information with time or other intervening activities. In the short term, however, massed practice is often better. Thus, whether one wishes to learn how to type, play badminton, speak a foreign language, or solve geometry problems, one should consider implementing a distributed practice schedule, even if such a schedule might induce more errors during practice or acquisition.

Variability of Practice

Similar to distributing practice, varying the conditions of practice or study sessions—for example, by having a trainee practice skills related to but different from the target skill—can also have detrimental effects on performance during acquisition but then foster long-term learning and transfer. Most of the research in this vein has focused on motor learning, although a handful of studies on verbal learning have also demonstrated the long-term benefits of practice variability. We now review research from both the motor- and verbal-learning traditions that has shown dissociable effects of practice variability on learning and performance.

Motor learning

Research on motor learning and practice variability suggests that if a basketball player, for example, wants to shoot accurate free throws, he or she should not only practice from the foul line itself but also from various positions neighboring the foul line. Such variable practice might not appear to be effective during practice—specifically, more performance errors would likely be induced relative to shooting only from the foul line—but would facilitate long-term learning. As discussed in more detail later, varying the conditions of practice seems to be effective for learning because it enables one to become familiar with, and learn to manipulate the parameters of, the general motor program underlying some skill, like shooting a basketball (Schmidt, 1975). We now discuss several findings from the motor-learning domain suggesting that increasing practice variability, while potentially inducing more errors during training, or acquisition, also has the potential to confer long-term learning benefits (for a review, see Guadagnoli & Lee, 2004).

In their important article, Kerr and Booth (1978) provided compelling evidence that varying the conditions of practice, as opposed to keeping them fixed, can boost long-term learning of a motor skill. In their study, children tossed beanbags at a target on the floor from distances of 2 and 4 feet (varied practice) or only 3 feet (fixed practice). After a delay, all participants were tested from a distance of 3 feet, the sole distance practiced by participants in the fixed-practice group. Intuition would suggest that participants in the fixed-practice group, who exclusively practiced from the tested distance, would do better than those in the varied-practice group, who never practiced at the tested distance. The results, however, showed the opposite pattern: Varying the practice distances led to more accurate tosses from 3 feet away on the final test, showcasing the benefits of variable practice in producing transfer of a motor skill, a result that has been replicated and extended (e.g., Pigott & Shapiro, 1984; Roller, Cohen, Kimball, & Bloomberg, 2001; Wulf, 1991).

Subsequent research found that variable practice can foster the learning of other complex motor skills, such as shooting a basketball (Landin, Hebert, & Fairweather, 1993) and mastering a forehand racket skill (Green, Whitehead, & Sugden, 1995). In the basketball study, two groups of participants practiced shooting basketball free throws over a period of 3 days. In the fixed-practice condition, participants shot the free throws exclusively from the criterion distance of 12 feet, whereas participants in the variable-practice condition shot from the criterion distance as well as from two other distances (8 feet and 15 feet). It is important to note that the total number of free throws (120) practiced by both groups was equated. The retention test, administered 72 hr after the practice phase, consisted of participants shooting 10 free throws from the criterion distance (12 feet). Again, the common-sense prediction would be that participants in the fixed-practice condition would make more free throws on the final test because they practiced more free throws from that distance compared with participants in the variable-practice condition. The counterintuitive finding, however, was that participants in the variable-practice condition made more free throws on the delayed-retention test, suggesting that practicing from multiple locations engendered more familiarity with the general motor program underlying the skill.

The learning of simpler motor skills has also been repeatedly shown to benefit from varying the conditions of practice, even in cases when such practice has detrimental effects on acquisition performance. Many of the studies that have produced this learning–performance interaction effect have examined timing skills (e.g., Catalano & Kleiner, 1984; Hall & Magill, 1995; Lee, Magill, & Weeks, 1985; Wrisberg & Mead, 1983; Wulf & Schmidt, 1988). For example, in one study, participants attempted to knock over a barrier with their arm from a given starting point, with the goal of doing so in precisely 200 ms. A variable group practiced from four different starting points (15, 35, 60, and 65 cm), whereas a constant group always practiced from the same starting point (e.g., 60 cm). As displayed in Figure 3, the variable group performed worse than the constant group during acquisition, producing more absolute errors when attempting to execute the arm movement in the target time of 200 ms, yet showed better learning on subsequent immediate and delayed (1 day) transfer tests in which a new starting point (50 cm) was tested (McCracken & Stelmach, 1977). A similar learning–performance interaction was shown in a study that examined the effects of variable practice in learning a criterion handgrip force. Compared with those who practiced solely to reach the criterion force, participants who practiced additional handgrip forces performed worse during acquisition at reaching the criterion force but were more accurate in producing the criterion force after a delay (C. H. Shea & Kohl, 1991; see also C. H. Shea & Kohl, 1990).

Verbal learning

One long-standing and widespread piece of advice regularly given to students is to find a quiet location—say, a favorite corner of the library—and to study there on a consistent basis. Keeping study conditions constant, it is thought, benefits learning. However, analogous to findings in the motor domain, studies have shown that inducing variation during study sessions—for example, by varying the environmental context in which to-be-remembered material is studied or increasing the variation of to-be-solved problems—can also benefit verbal learning. Inducing such variation often has negligible effects on acquisition performance, or may even impede

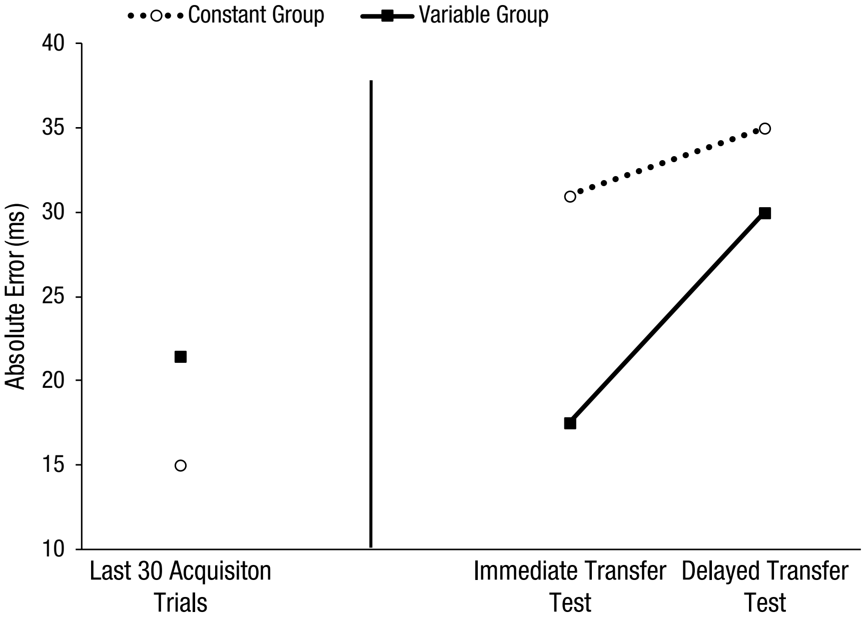

Fig. 3. Absolute timing errors in performing a ballistic timing task during the acquisition phase and on immediate and delayed transfer tests as a function of practice condition (constant vs. variable). (Note that lower scores represent better performance and learning.). Data are adapted and approximated from McCracken and Stelmach (1977).

it, but it can enhance long-term learning because the material becomes associated with a greater range of memory cues that serve to facilitate access to that material later. Several studies in the verbal-learning tradition have demonstrated this empirically.

One study examined whether varying the physical context, or environment, in which material is studied can bolster learning when that material is tested in a new context, an issue that remains relevant given that modern standardized tests (e.g., SAT, GRE) are often administered in unfamiliar locations. The participants first studied a list of 40 words. Half the participants studied the list in Room A, a particular location on the University of Michigan campus; the other participants studied the list in Room B, a different location on the Michigan campus. Three hours later, half of the participants in each group restudied the words again in the same room, whereas the other participants studied the list again in the other location. On the final test, administered 3 hr after the second study session, all participants were tested on the words in a neutral location, Room C. Strikingly, participants who studied in different rooms recalled approximately 21% more of the words than participants who studied in the same room, demonstrating the mnemonic benefit of variable practice (Smith, Glenberg, & Bjork, 1978). Thus, if participants were tested in a novel location, varying the physical study environments bolstered learning, a finding that was later replicated using the same or similar materials (Glenberg, 1979; Smith, 1982). Another study replicated this finding with more complex learning material by showing that participants' 5-day retention of statistical concepts was better when it occurred after four successive lectures given in four different locations as opposed to when all of the lectures were given in the same location. Performance on short-term retention tests administered immediately after each statistics lecture, however, was similar for both groups (Smith & Rothkopf, 1984).

In addition to increasing the variation of the environmental context, long-term learning and transfer but not necessarily short-term performance can also profit from increasing the variation of problems during an acquisition phase. For example, in a study that examined the effects of variable practice on a task that involved troubleshooting a computer-based simulation of a chemical process plant, participants produced a pattern of results indicative of a "transfer paradox." Specifically, highly variable practice problems, relative to low-variability problems, induced more performance errors during practice but had positive effects on learning, as evidenced by the number of new problems solved on a later test (Van Merrienboer, de Croock, & Jelsma, 1997). Such encoding variability has also been shown to enhance analogical reasoning (Gick & Holyoak, 1983) and geometrical problem solving (Paas & Van Merrienboer, 1994), as well as the retention of text material (Mannes & Kintsch, 1987) and face–name pairs (Smith & Handy, 2014).

In yet another study that demonstrated learning benefits of variable practice with verbal materials, participants practiced solving anagrams by either repeatedly solving the anagram that was tested later (e.g., *LDOOF* was solved three times during the practice phase and appeared on the test) or solving multiple versions of the anagram that was tested later (e.g., *DOLOF, FOLOD*, and *OOFLD* were practiced and *LDOOF* appeared on the test). Of interest was whether solving multiple variants of the anagram—that is, increasing the variability of the problems—would enhance participants' ability to solve the anagrams later. Indeed, despite participants in the variable practice condition taking relatively longer to solve the anagrams during the practice phase, revealing a short-term performance decrement, they solved relatively more of the anagrams on a later test (M. K. Goode, Geraci, & Roediger, 2008). Like several of the results reviewed in the previous section on motor learning, this result is counterintuitive because the variable practice group never attempted to solve the specific anagram that was later tested, whereas the other group solved it three times during the practice phase. Thus, these results conceptually replicate the outcome of Kerr and Booth's (1978) motor-learning experiment, in which tossing beanbags at a target from various nontested distances was better for learning than practicing those tosses from the tested distance.

Summary

The long-term retention and transfer of motor skills—both simple and complex—often profit from the type of practice that entails one to perform multiple iterations, rather than a single iteration, of those motor skills, despite such practice potentially having negligible or even negative effects on performance during training. The same has been revealed in verbal-learning experiments that have increased the variation of to-be-solved problems or varied the environmental context in which to-be-remembered material was studied. Variable practice, it seems, broadens one's familiarity with the general underlying motor skill or knowledge base needed to successfully perform a task.

Retrieval Practice

Decades of research suggest that the retrieval processes triggered by testing actually changes the retrieved information in important ways. That is, tests act not only as passive assessments of what is stored in memory (as is often the traditional perspective in education) but also as vehicles that modify what is stored in memory. This section reviews evidence from both the motor- and verbal-learning domains that lead to such a conclusion. In the motor-skills literature, for example, to-be-learned movements that are self-produced are typically better learned than those that are externally guided or simply observed. Likewise, testing one's memory for verbal information, or having participants generate the information themselves, enhances long-term retention of that material compared with reading it over and over, even in cases when corrective feedback is not provided. A critical finding, relevant to the learning–performance distinction, is that conditions of retrieval practice that often facilitate long-term retention frequently may appear unhelpful in the short term compared with their counterpart conditions.

Motor learning

When teaching a motor skill, such as a gymnastics flip or a golf swing, it is commonplace for instructors to physically guide the learner through the desired motions. Intuition suggests that this type of instruction should be beneficial; indeed, research has shown that guiding learners reduces performance errors during acquisition compared with when learners attempt to produce the skill without guidance (i.e., are encouraged to retrieve the skill on their own). The problem is that on assessments of long-term learning when guidance can no longer be relied on, the reverse is often true—that is, practicing a skill without guidance frequently produces better learning than does being guided during acquisition (for a review on guidance research in motor-related tasks, see Hodges & Campagnaro, 2012). The long-term learning of motor skills, but not necessarily short-term performance, also profits from a test (as opposed to a restudy opportunity) and when learners are permitted to generate their own to-be-remembered motor skills (as opposed to when the skills are chosen for them).

Early research on the effects of guidance (e.g., Melcher, 1934; Waters, 1930) showed that providing physical assistance during the acquisition of simple to-be-learned movements had positive effects when participants were subsequently asked to perform those movements on their own, suggesting that learning profits from initial guidance. However, the retention intervals in these studies were particularly short, and thus any claims of long-term learning were tenuous. It was not until decades later that the first studies to examine the long-term effects of guidance emerged. In one such study, participants practiced a joystick pursuit-tracking task while either being physically guided by another person or not. The guided group outperformed the unguided group during training and on initial short-term performance tests, but on a later retention test administered 6 weeks later, the unguided group demonstrated better learning than the guided group. Furthermore, the guided group failed to show better retention than a group of participants who had never performed the task but simply watched (Baker, 1968; see also Armstrong, 1970).

Subsequent research has replicated and extended the learning and performance effects of guidance. For example, on a task that involved manipulating a lever to various positions, a physically guided group performed better during acquisition (i.e., made fewer performance errors) but worse after a retention interval, relative to an unguided group (Winstein, Pohl, & Lewthwaite, 1994). Likewise, during training of a bimanual coordination task that involved arm extensions, guided practice prevented performance errors; however, it also yielded less long-term learning compared with conditions in which partial guidance or no guidance was provided (Feijen, Hodges, & Beek, 2010; see also Tsutsui & Imanaka, 2003). Finally, in a study that examined whether a harness could serve as an aid to properly modify the bowling technique involved in the sport of cricket, it was found that the restriction applied by the harness improved techniques in the short term but failed to yield any long-term learning benefits, compared with when no harness was used (Wallis, Elliot, & Koh, 2002). Clearly, guidance during training can have differential effects on learning and performance.

Another, rather simple way to examine the effects of retrieval practice on learning and performance is to allow learners to first observe the to-be-learned skill and then either test the learners (i.e., require them to reproduce, or retrieve, the skill on their own) or present the skill again without the requirement to reproduce it. A subsequently administered test could then reveal whether retrieval practice, relative to re-presentation trials, enhances learning. It is surprising that scant empirical work in the motor-learning domain has used this sort of method to better understand the potential benefits of retrieval practice. Representing a notable exception, one study examined the effects of retrieval practice on learning an arm-positioning task. After an initial presentation of to-be-learned positions, participants either were tested on the positions several times or were simply re-presented with them over and over without being tested. Participants who engaged in retrieval practice showed better long-term retention of the arm positions than those who simply observed the positions multiple times. The opposite was true, however, when performance was assessed during acquisition (Hagman, 1983). Subsequent motor-skills research replicated the long-term learning benefits conferred by this type of retrieval practice (i.e., testing vs. restudying; Boutin et al., 2012; Boutin, Panzer, & Blandin, 2013).

Finally, the learning of motor skills profits from another form of retrieval practice—namely, permitting learners to generate their own to-be-learned movements as opposed to the movements being selected for them. In one of the earliest and most convincing demonstrations of this *preselection effect*, participants reproduced rapid arm movements that were either previously selected by themselves or imposed by the experimenter. Retention of the arm movements—in terms of both rapidity and precision—favored the selection group, even though no indicators of such long-term learning could be gleaned from the acquisition phase (Stelmach, Kelso, & Wallace, 1975). The preselection effect quickly emerged as one of the most robust and reliable effects in the motor-learning literature (see also Martenuik, 1973; for an early review, see Kelso & Wallace, 1978).

Verbal learning

Similar to the research in the motor-learning domain, empirical work investigating retrieval practice (or testing) of verbal material dates back decades, out of which has emerged the consensus that retrieving information from memory does more than simply reveal that the information exists in memory. In fact, the act of retrieval is a "memory modifier" (R. A. Bjork, 1975) in the sense that it renders the successfully retrieved information more recallable in the future than it would have been otherwise, a finding that has been termed the *testing effect*, which has been demonstrated across the life span using a wide range of materials and outcome measures (for reviews, see Carpenter, 2012; Roediger & Butler, 2011; Roediger & Karpicke, 2006a). In other words, retrieval practice is itself a potent learning event. In the short term, however, retrieval practice often appears to fail to confer any mnemonic benefits compared with conditions in which the material is restudied instead of tested. We now consider work on retrieval practice in the verbal-learning domain that has necessitated the distinction between learning and performance.

Although the first large-scale studies on the testing effect can be traced back to Gates (1917) and Spitzer (1939), it was not until the 1970s that researchers provided compelling evidence that retrieval practice can have differential effects on learning and performance. In one study, for example, participants studied 40 single words either three times before taking a free-recall test (SSST) or once before taking three free-recall tests (STTT). During the fourth phase of this procedure in which both groups were tested, participants assigned to the SSST condition showed greater short-term recall performance than those in the STTT condition—in other words, repeated studying was better than repeated testing. Long-term recall assessed 2 days later, however, favored the STTT condition (Hogan & Kintsch, 1971). These findings were later replicated and extended in a study that found that repeated studying led to better recall than repeated testing after 5 min (50% vs. 28%) but that repeated testing trumped repeated studying, albeit only slightly, on a delayed recall test administered 2 days later (25% vs.

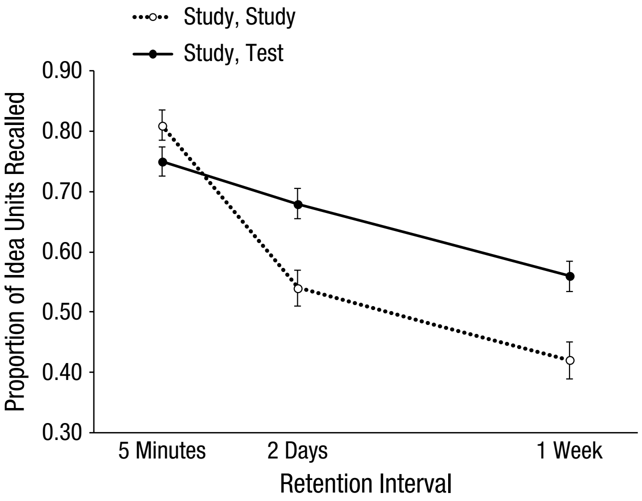

Fig. 4. Proportion of idea units correctly recalled on immediate (5 min) and delayed (2 days and 1 week) retention tests after participants studied the passages either twice or once before taking an initial test. (Note that higher scores represent better performance and learning.) Error bars represent standard errors of the means. Data are adapted from Roediger and Karpicke (2006b).

23%; Thompson, Wenger, & Bartling, 1978). That same year, expanded-interval testing schedules were found to produce better recall of to-be-learned names than equal-interval testing schedules, but both of these conditions led to better long-term learning than did a massed-testing condition, in which several tests were administered in succession immediately after the presentation of a given name, a condition that showed nearly errorless performance during the acquisition phase (Landauer & Bjork, 1978).

In another study that provided a convincing demonstration of a learning-performance interaction as it relates to retrieval practice, one group of participants (repeated study) studied a 40-word list five consecutive times (SSSSS), whereas another group (repeated test) studied the list once before four consecutive recall tests (STTTT). Final recall tests were then administered to different groups of participants (from each group) after 5 min or 1 week. The repeated-study group outperformed the repeated-test group by a large margin on the immediate (5 min) test, but on the delayed (1 week) test, the opposite pattern was observed—specifically, repeated testing led to better long-term retention than did repeated studying. It was also clear that testing helped stabilize memory, as forgetting over time was far more pronounced in the repeated-study group than the repeated-test group. When specifically considering 1-week recall as a percentage of 5-min recall, researchers found that repeated studying and repeated testing were associated with approximately 75% and 30% forgetting, respectively (Wheeler, Ewers, & Buonanno, 2003).

Thus far, we have reviewed studies on the testing effect that have used relatively simple learning materials (e.g., single words, word pairs); however, it is also clear that retrieval practice can have differential effects on learning and performance when more educationally relevant materials are used. One such study involved participants first studying prose passages covering general topics, such as the sun and sea otters. In one condition, participants then restudied the passage in its entirety, whereas in another condition, participants were tested, without feedback, for their ability to recall the studied material. Final recall tests were then administered to different groups of participants from each condition after 5 min, 2 days, or 1 week. The results, which are shown in Figure 4, are clear. After 5 min, participants who restudied the passage showed better recall performance than did participants who took an intervening test without feedback. On the delayed retention tests, however, there was a significant reversal such that the tested group recalled more of the material after 2 days and 1 week than the restudy group, a finding that was subsequently replicated and extended in the study's second experiment (Roediger & Karpicke, 2006b). What makes the results of this study (and others like it) particularly impressive is that no feedback was given to participants in the tested condition during the initial test, which means that participants in the test condition were reexposed only to the material they were initially able to recall—approximately 70% of the passage—whereas participants in the restudy condition were reexposed to the entire passage before the final retention tests. Despite this disadvantage, participants in the tested group retained more information over the long term.

Research on the *generation effect*, a closely related phenomenon to the testing effect, also points to the long-term learning benefits of retrieval practice (for important differences between the generation effect and the testing effect, see Karpicke & Zaromb, 2010). In a typical generation experiment, participants are asked to either generate the to-be-learned items themselves—for example, by producing opposites when presented with a word (e.g., *hot–???*)—or to simply read the items (e.g., *long–short*). A later retention test is then administered, which usually consists of presenting the cues (*hot–???, long–???*) and asking participants to recall their corresponding targets (*cold, short*). Slamecka and Graf (1978; see also Jacoby, 1978) are often credited as the first to demonstrate that generating items from semantic memory is better for learning than simply reading them, a finding that has been replicated hundreds of times using various materials, procedures, and outcome measures (for a review, see Bertsch, Pesta, Wiscott, & McDaniel, 2007). For current purposes, it is important to note that unless participants can successfully generate every to-be-generated item during the study phase (which almost never happens), generated items will always be associated with worse acquisition performance than read items if a test was given immediately after each item. This is because, similar to unsuccessful retrieval attempts in testing-effect studies, unsuccessful generation attempts prevent exposure to the material that will be tested later. Despite this short-term performance hindrance, generation still enhances long-term learning.

Even more compelling evidence in favor of the learning–performance distinction comes from research that has revealed that learning can profit from generation attempts that are assured to be incorrect during acquisition, a phenomenon that was demonstrated some time ago (Kane & Anderson, 1978; Slamecka & Fevreiski, 1983) and is now garnering considerable empirical attention once again (Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013; Huelser & Metcalfe, 2012; Knight, Ball, Brewer, DeWitt, & Marsh, 2012; Kornell, Hays, & Bjork, 2009; Potts & Shanks, 2014; Yan, Yu, Garcia, & Bjork, 2014).

This resurgence in interest in the potential benefits of failed generation was spurred by research using a paradigm in which participants study weakly related word pairs, some of which are presented intact (e.g., *whale–mammal*) for study, whereas the others require that the participants, on the basis of the cue by itself (e.g., *whale–???*), first try to predict the to-be-learned response. Critically, by choosing weakly related pairs as the materials, experimenters can ensure that participants almost always fail to guess the correct target. That is, when presented with *whale*, participants will almost always generate something other than *mammal* (e.g., *big, ocean, blue*). Nevertheless, across multiple experimental designs using this paradigm, failed retrieval attempts prior to encoding were found to enhance learning (Kornell et al., 2009). One possible explanation for this effect is that attempting to predict the to-be-learned response activates the broad semantic network associated with the cue word, which, in turn, may facilitate associating the response to the cue (Grimaldi & Karpicke, 2012; Hays et al., 2013; Huelser & Metcalfe, 2012; but see Potts & Shanks, 2014). More generally, this research indicates—as counterintuitive as it may seem—that the production of errors during acquisition can, under some circumstances, actually boost long-term retention.

Summary

Evidence from both the motor- and verbal-learning domains shows that retrieval practice can have opposing effects on learning and performance. Motor-learning studies have revealed that, on the whole, physical guidance often reduces performance errors during training but that unguided, active involvement promotes better long-term retention of skills. Likewise, practicing retrieval of verbal materials may appear unhelpful during acquisition and on immediate memory tests, but it provides substantial benefits in preserving or stabilizing long-term memory. It would seem prudent, therefore, that trainers and instructors incorporate retrieval practice into their curriculum and that students test themselves as a means to optimize their own learning.

Metacognition

To what extent are educators and students aware of what activities are beneficial for long-lasting learning? In particular, what does a learner need to know to manage his or her own self-regulated learning in an optimal way? These important questions concern *metacognition*, which, broadly construed, refers to thinking about thinking (see Nelson, 1996). In the domain of learning and memory, it denotes more specifically (a) one's knowledge and understanding of how learning and memory operate and (b) the interplay between the monitoring and controlling of one's own ongoing learning and memory or that of others (for reviews, see R. A. Bjork, Dunlosky, & Kornell, 2013; Soderstrom, Yue, & Bjork, in press). Elucidating how people think about and monitor their own learning is paramount because subjective experience plays a causal role in determining subsequent behavior (e.g., deciding what material should be restudied and for what duration), and thus the appropriateness of such behavior will necessarily depend on the meaning and validity of learners' subjective experiences (see Nelson & Narens, 1990).

Although there is overwhelming empirical evidence that learning and performance are dissociable, there appears to be a lack of understanding on the part of instructors and learners alike that performance during acquisition is a highly imperfect index of long-term learning. As a consequence, what is effective for learning is often misaligned with our metacognitive assessments of what we think is effective for learning (for a review, see R. A. Bjork, 1999). This disconnect has been clearly demonstrated in surveys of students' beliefs about learning. For example, one study investigated undergraduates' awareness of six empirically supported learning strategies, three of which—spacing versus massing, testing versus restudying, and generating versus reading—we have discussed in earlier sections of this review. Overwhelmingly, students endorsed as most effective those strategies known to enhance short-term performance, a pattern that was strikingly evident when students were confronted with choosing between massed or spaced study: 93.33% of surveyed students incorrectly endorsed massed study as being more effective for learning than spaced study (J. McCabe, 2011). From a research perspective, this is quite remarkable (and alarming) considering that the spacing effect has been demonstrated hundreds of times in the past century and has emerged as one of the most robust and reliable effects in all of memory research. Fortunately, this same study also found that educational interventions—for example, a cognition course or targeted instruction on effective learning techniques—helped ameliorate these misconceptions.

Other surveys have investigated how students study on their own. For example, a survey of 472 college students found that most students reported using a rereading strategy. Additionally, although 90% of students reported using self-testing, most them reported doing so in order to identify gaps in their knowledge, rather than because they believed that self-testing conferred a direct learning benefit. Moreover, 64% of students reported not revisiting material once they felt like they had mastered it, while only 36% of students reported that they would restudy or test themselves later on that information (Kornell & Bjork, 2007; for similar results, see Hartwig & Dunlosky, 2012). Another survey showed that students are generally unaware of the benefits of retrieval practice compared with rereading. When asked to report and rank their own study strategies, 84% of students ranked rereading as one of their strategies of choice, whereas only 11% of students reported using retrieval practice at all (Karpicke, Butler, & Roediger, 2009). It was argued that students prefer rereading because it produces a heightened sense of fluency or familiarity with the material, which students misinterpret as an index of learning. In other words, students seem to favor rereading because it leads to relatively greater perceived gains in performance.

Similar illusions of competence have been demonstrated in research that has examined how and to what degree of accuracy people can monitor or evaluate their own learning. Such experimental research has used both retrospective and prospective judgments. With respect to retrospective judgments—that is, subjective evaluations of learning that require the learner to assess some past experience—people often erroneously endorse relatively ineffective conditions of learning. In Baddeley and Longman's (1978) study involving British postal workers, for example, distributing practice was better than massing it for the long-term retention of data entry (i.e., keystroke) skills; however, learners in the distributed-practice group reported being relatively less satisfied with their training because they felt they were falling behind the massed-practice group, which, in fact, was true during the acquisition phase. Thus, learners appear to interpret short-term performance as a reliable guide to long-term learning.

Such biased retrospective judgments have also been shown after tests of inductive learning. In one study, for example, participants learned artists' painting styles according to a study schedule that was either massed (blocked)—that is, every painting from an artist was presented successively before moving on to the paintings from a new artist—or spaced (interleaved)—that is, paintings from several artists were mixed together. On a final induction test, participants were presented with new paintings and were asked to identify which of the previously studied artists painted them. Such a test is considered a test of inductive learning because success on such a test requires one to have extracted the artists' general painting styles from sets of exemplars. The results clearly showed that inductive learning was enhanced by the spaced study schedule compared with the massed schedule. It is interesting to note, however, that when asked after the induction test which study schedule helped them learn better, an overwhelming majority of participants endorsed massing (Kornell & Bjork, 2008). This finding is all the more remarkable given that participants had already experienced the test in which their learning profited from interleaving. Subsequent research on inductive learning has demonstrated that learners, when permitted to choose their own study schedule, also prefer massing (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013). Thus, not only does massing produce a sense of fluency during acquisition performance that is misinterpreted as learning, but learners also seem to hold the misguided theory that massing one's study is an effective way to learn.

Prospective metacognitive judgments, like retrospective ones, can also be heavily influenced by short-term performance. For current purposes, the most relevant prospective judgment—and the one that garners the most empirical attention in contemporary metacognitive research—is the judgment of learning (JOL), which is typically solicited during an acquisition, or encoding, phase. Here, learners are asked to predict—usually on a 0%–100% scale—the likelihood that some information will be remembered later. In other words, JOLs involve learners predicting their own learning. Collecting such predictions permits an examination of how learners decide which information has been learned and which has not and how well those predictions correspond to actual learning on a later test. Although some early work on verbal learning showed that JOLs predicted actual learning relatively well (e.g., Arbuckle & Cuddy, 1969), the more recent JOL literature in this domain is rife with examples in which people's immediate JOLs are not diagnostic of future learning (e.g., Benjamin, Bjork, & Schwartz, 1998; Castel, McCabe, & Roediger, 2007; Koriat & Bjork, 2005; Koriat, Bjork, Sheffer, & Bar, 2004; Mazzoni & Nelson, 1995; Rhodes & Castel, 2008; Roediger & Karpicke, 2006b; Soderstrom & McCabe, 2011; Yue, Castel, & Bjork, 2013), revealing striking illusions of competence and compelling evidence that JOLs are inferential in nature, based on cues rather than on memory

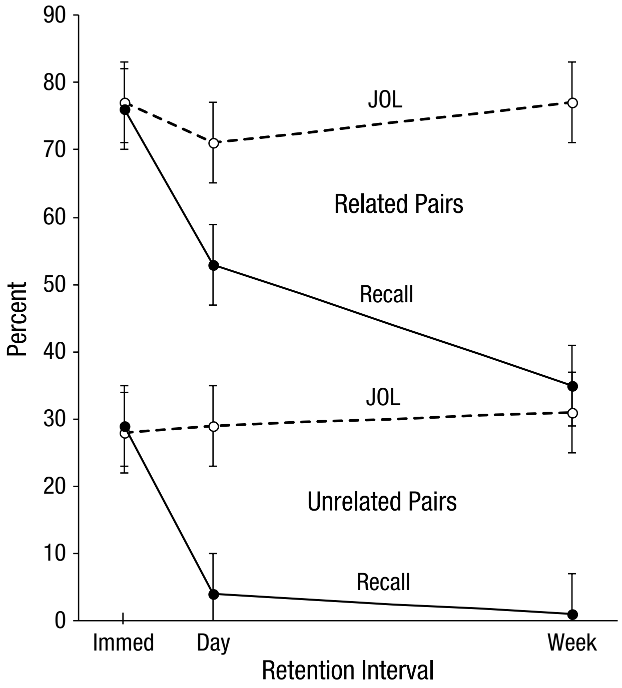

Fig. 5. Mean judgment of learning (JOL) and recall as a function of retention interval (immediate [Immed], 1 day, or 1 week) for related and unrelated word pairs. (Note that higher scores represent elevated predictions [JOLs] and better performance and learning [recall].) Error bars represent 95% confidence intervals. Data are adapted from Koriat, Bjork, Sheffer, and Bar (2004).

strength (see Koriat, 1997). We now discuss several examples in which learners' JOLs have exposed misconceptions about learning (for a more comprehensive review, see Schwartz & Efklides, 2012).

Generally speaking, learners tend to be overconfident in predicting their own learning and regularly exhibit what has been termed the *stability bias*, which refers to the tendency to believe that current accessibility of retrieved information (i.e., performance) will remain stable across time, rather than appreciating those factors that may impair or enhance later learning (Kornell & Bjork, 2009; see also Ariel, Hines, & Hertzog, 2014; Kornell, 2011). In a study that demonstrated a particularly striking example of a stability bias, participants studied related and unrelated word pairs, making JOLs after each item. Separate groups of participants were asked to base their predictions on how well they would remember the pairs on an immediate test, a test after 1 day, or a test after 1 week. As illustrated in Figure 5, participants produced a pattern of results demonstrating apparent insensitivity to retention interval—specifically, equivalent JOLs were given across the three retention intervals. However, and as expected, actual recall decreased as a function of retention interval. Also evident in Figure 5 is that JOLs were highly sensitivity to the relatedness of the word pairs, which led the authors to conclude that encoding fluency, or how easily information is processed during study, can largely drive JOLs (Koriat et al., 2004; see also Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Koriat, 2008; Undorf & Erdfelder, 2011; for an alternative account, see Mueller, Tauber, & Dunlosky, 2013), even at the expense of other extremely relevant information—in this case, retention interval.

Other research supports the conjecture that *retrieval* fluency (i.e., how easily to-be-remembered information is retrieved during an acquisition phase), like encoding fluency, can also influence JOL magnitude (see, e.g., Benjamin et al., 1998; Hertzog, Dunlosky, Robinson, & Kidder, 2003). In one demonstration of this, participants answered relatively easy general-knowledge questions, after which time they predicted, on an item-by-item basis, the likelihood that they would be able to recall a given answer on a later free-recall test—that is, the likelihood they would be able to recall having given a particular answer without the question being provided again. The results indicated that answers that took the shortest time to generate were given higher JOLs compared with those answers that were generated slowly. In other words, participants based their JOLs on short-term performance—in this case, retrieval latency. Later recall, however, showed the opposite pattern: Answers that took a longer time to generate were recalled at a higher rate than were answers generated more quickly, presumably because the effort involved in generating an answer is positively related with its subsequent recall (Benjamin et al., 1998). Thus, while retrieval fluency was related to both JOLs and later recall, the direction of this relationship differed whether it was assessed subjectively (via JOLs) or objectively (via final recall), demonstrating, among other things, that learners are captured by gains in short-term performance and can mistakenly conflate such gains with long-term learning.

The benefits of retrieval, more generally, are not appreciated by learners either (see, e.g., Karpicke, 2009; Kornell & Son, 2009). As discussed previously, the testing effect refers to the finding that retrieval practice acts as a learning event, rendering retrieved information more recallable in the future than it would have been otherwise (see Roediger & Karpicke, 2006a). In Roediger and Karpicke's (2006b) study in which participants studied prose passages and then were either tested on those passages or restudied them, long-term retention, measured 1 week after the study phase, increased as a function of testing opportunities during acquisition. However, participants predicted the opposite pattern—specifically, that learning after 1 week would be best when the passages were studied multiple times without being tested, a pattern of performance that was, in fact, demonstrated in the short term (after 5 min). Again, learners seem to assume that whatever boosts performance will also profit long-term retention.

Finally, as is the case with retrospective judgments, higher JOLs are given to material or skills that are studied or practiced in a massed (blocked) schedule compared with a spaced (distributed) schedule. In one example, participants were presented with a list of to-be-remembered words. Within the list, a second repetition of each item occurred either immediately after its first presentation (massed) or following a number of other items (spaced). Participants predicted that the massed items would be better remembered than the spaced items, whereas actual recall showed the opposite pattern (Zechmeister & Shaughnessy, 1980). An analogous result was subsequently found for the learning of simple keystroke patterns: Despite the fact that distributed practice led to relatively greater long-term gains in learning the keystrokes, participants predicted that blocked practice—which did boost short-term performance—would be better after a delay (Simon & Bjork, 2001). This mismatch between JOLs and actual learning—that people's JOLs favor massed practice, whereas actual learning profits more from distributed practice—has also been replicated for the learning of piano melodies (Abushanab & Bishara, 2013).

Given that people are generally unaware of what activities are beneficial for long-term retention and that learners, by and large, have trouble accurately monitoring their own ongoing learning, it is important to identify ways to foster metacognitive sophistication in order to optimize self-regulated learning (see R. A. Bjork et al., 2013). Instructors and students, for example, need to become familiar with the types of learning strategies that promote long-term learning—some of which we have already discussed in the present review—before we can expect the use of such strategies to be encouraged by teachers or adopted by their pupils (for a review of the utility of various study strategies, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). As well, research in metacognition should endeavor to find methods of improving people's monitoring capabilities such that learners become accurate forecasters and, as a result, effective managers, of their own ongoing learning. Fortunately, the number of studies on this topic is mounting (e.g., Castel, 2008; DeWinstanley & Bjork, 2004; Koriat & Bjork, 2006; D. P. McCabe & Soderstrom, 2011; Nelson & Dunlosky, 1991; Soderstrom & Bjork, 2014; Soderstrom & Rhodes, 2014; Thiede & Anderson, 2003; Tullis, Finley, & Benjamin, 2013), a trend that we hope continues given the importance of such work.

Summary

Both survey and experimental research in metacognition have revealed that learners often mistakenly conflate short-term performance with long-term learning, ostensibly thinking, "If it's helping me now, it will help me later." The extant survey literature on beliefs about learning suggests that students, by and large, endorse and use strategies that may confer short-term performance gains but do not foster long-term learning. Likewise, research that has examined how people monitor their own ongoing learning has revealed that both retrospective and prospective judgments are heavily influenced by acquisition factors, a bias that often produces striking illusions of competence. Given that people act on their subjective experiences, it is imperative that people learn how to learn by becoming knowledgeable of what effective learning entails. It is important, too, that such metacognitive sophistication is fostered early on in one's education.

Contemporary Theoretical Perspectives

As discussed earlier, learning theorists from decades ago (e.g., Estes, 1955a; Guthrie, 1952; Hull, 1943; Skinner, 1938; Tolman, 1932; Tulving & Pearlstone, 1966) used terms in their own theories that distinguished between learning and performance. Given the early research on latent learning, overlearning, and fatigue, this distinction was necessary. To account for more recent empirical work in the motor- and verbal-learning domains, contemporary learning theorists also differentiate between the relatively permanent changes in behavior and knowledge that characterize long-term learning and the temporary fluctuations in performance that occur across the training or acquisition process. Although we briefly mentioned possible explanations of several of the various empirical findings reported in previous sections of this review, we now discuss in more detail the dominant contemporary learning theories that address the distinction between learning and performance.

R. A. Bjork and Bjork (1992), in an attempt to formulate an account of a wide range of fundamental human learning phenomena, resurrected the learning–performance distinction in their new theory of disuse by introducing the terms *storage strength* and *retrieval strength*. Storage strength refers to the degree to which memory representations (i.e., knowledge and procedures) are integrated or entrenched with other memory representations, whereas retrieval strength represents the current ease of access or activation of those memory representations given current mental and environmental cues. Current performance, which can be observed, is indexed by retrieval strength, whereas long-term learning is indexed by storage strength, which acts as a latent variable by enhancing the gain of retrieval strength during opportunities for study or practice and impeding the loss of retrieval strength across time and intervening or interfering events. Furthermore, storage capacity, unlike retrieval capacity, is assumed to be limitless and, once accumulated, never lost. This latter assumption—that the storage strength of memories are permanent—distinguishes Bjork and Bjork's new theory of disuse from Thorndike's (1914) original law of disuse, which asserted that memories, without continued use, will decay over time and can eventually disappear entirely.

According to the new theory of disuse, gains in storage strength are expressed as a negatively accelerated function of current retrieval strength—that is, the more accessible representations are in memory when study or test events occur, the less gains in storage strength can be achieved for those representations. Consequently, conditions that increase current retrieval strength might benefit performance in the short term but will fail to produce the type of permanent changes that characterize long-term learning. In contrast, situations that reduce current retrieval strength (i.e., produce forgetting)—for example, distributing study or practice sessions (as opposed to massing them), varying the conditions of learning (as opposed to keeping them constant), and encouraging retrieval practice (as opposed to restudy)—yield relatively greater gains in storage strength and thus lead to enhanced long-term retention and transfer. As argued by R. A. Bjork (2011), this interplay between retrieval strength and storage strength—namely, that forgetting can foster learning—is adaptive, yet counterintuitive, and has broad implications for treatment (see R. A. Bjork & Bjork, 2006; Lang, Craske, & Bjork, 1999) as well as training.

From a formal-modeling standpoint, the new theory of disuse shares a number of properties with contextual-fluctuation models (see, e.g., Mensink & Raaijmakers, 1988, whose model traces back to the influential stimulus-fluctuation model proposed by Estes, 1955a, 1955b). The basic idea is that the performing–learning organism is heavily influenced by current cues, which gradually change or fluctuate as time and events go on and different aspects of the external and internal environments are "sampled." When cues are not changing, or are changing slowly, as in massed practice, for example, performance will increase rapidly, but forgetting will be rapid as well, as cues change across a retention interval. As contextual variation across acquisition trials either is introduced or occurs spontaneously, performance will improve more slowly, but more total cues will become associated with to-be-learned responses, which will enhance learning, as measured after a delay or in an altered context. Basically, to borrow Estes's initial language, response strength (performance) is indexed by how associated some target response is to the current cues, whereas habit strength (learning) is indexed by how much some target response is associated to the whole range of cues that characterize some task and situation.

In the motor skills literature, specifically, the schema theory of motor control and the reloading hypothesis offer highly cited explanations for the learning and performance effects produced by variable and distributed practice, respectively. Originally postulated by Schmidt (1975), the schema theory of motor control claims that variable practice—that is, practicing iterations of a skill that are related to but different from the target skill—fosters long-term learning because it sensitizes one to the general motor program, or schema, underlying a skill (see also Schmidt, 2003). To flesh out this notion, consider that discrete motor skills (e.g., shooting a basketball, serving a tennis ball, swinging a golf club) involve the coordination and implementation of classes of simpler movements, each associated with unique parameters, such as its timing, speed, and force. In order to successfully reconstruct the parameters of the movements required to execute a given skill, learners need to become familiar with how the various rules that govern one class of movements are related to the rules that govern the other relevant classes of movements and how such interdependencies affect outcomes. An effective way of doing this, according to schema theory, is to increase the variation of the practiced skill such that one is required to learn how to adjust the necessary movement parameters to achieve desired goals. As we have already discussed in this review, practice variability, while having the potential to induce more errors during acquisition compared with fixed practice conditions, often leads to substantial gains in long-term retention and transfer.

In terms of the learning benefits associated with distributed practice, the reloading hypothesis asserts that spacing out practice sessions with time or other activities encourages the "reloading," or reproducing, of the motor programs needed to execute to-be-learned skills (Lee & Magill, 1983, 1985). This is because the spacing inserted between practice sessions results in a temporary loss of access to the relevant motor commands. The effortful processing required to reload the commands during distributed practice appears to facilitate learning but impede short-term performance, compared with blocked (massed) practice in which skills are performed over and over again.

Last, the general idea that what can hurt performance can help learning is captured in the desirable difficulties framework (R. A. Bjork, 1994; see also, E. L. Bjork & Bjork, 2011; R. A. Bjork, 2013). Manipulations such as distributed practice, variable practice, and retrieval practice are "desirable" because they support better long-term retention and transfer compared with their counterpart conditions. Such effective learning manipulations are also "difficult," however, in the sense that they can degrade performance during acquisition or training and, consequently, are likely to be interpreted as ineffective by instructors and students alike. As unintuitive as it may seem, the active cognitive processes engendered by confronting and resolving difficulties during acquisition serves to effectively link or entrench new information with knowledge that already exists in memory. Furthermore, given that these active processes are also likely recruited during later assessments of long-term retention, the notion of desirable difficulties generally accords with the principle of transfer-appropriate processing (Morris, Bransford, & Franks, 1977) and the related encoding specificity principle (Tulving & Thomson, 1973), both of which contend that memory will improve to the extent that the engaged study and test processes overlap. It is important to note, however, that when the difficulties cannot be overcome by the learner—for example, when previously encountered information cannot be successfully retrieved during retrieval practice—they become undesirable (see McDaniel & Butler, 2010). Thus, an ongoing challenge for researchers has been to identify when difficulties are desirable for learning and when they are not, so as to appropriately inform the instructional practices of instructors and the study behaviors of students.

Summary

Several current theoretical perspectives make the crucial distinction between short-term performance and long-term learning. According to the new theory of disuse, dissociable effects of learning and performance arise as a result of the adaptive interplay between storage strength—the extent to which new and prior knowledge is integrated—and retrieval strength—the relative ease with which information can be accessed. The schema theory of motor control claims that variable practice promotes long-term retention and transfer by familiarizing learners with the general motor programs that underlie motor skills. Likewise, the reloading hypothesis asserts that distributed practice encourages learners to reload or reproduce the to-be-learned motor skills during acquisition, which is a potent learning event, despite appearing not to be during acquisition. Finally, the desirable difficulties framework proposes that manipulations that appear to be difficult—both objectively and subjectively—during acquisition or training can be desirable for long-term retention and transfer because they engender active encoding processes.

Concluding Comments

We have provided the first integrative review of the overwhelming empirical evidence that necessitates the critical distinction between learning—the relatively permanent changes in behavior or knowledge that support long-term retention and transfer—and performance—the temporary fluctuations in behavior or knowledge that are observed and measured during training or instruction or immediately thereafter. Dating back nearly a century ago, early research on latent learning, overlearning, and fatigue provided the first insights into the learning–performance distinction by showing that substantial learning could occur in the absence of any discernible changes in performance. This work—conducted with both humans and nonhuman animals—compelled learning theorists at that time to make corresponding conceptual distinctions in their own theories of learning and memory. More recent research in the motor-skills and verbal-learning literatures have demonstrated the converse to also be true—specifically, that changes in short-term performance often bear no relationship to long-term learning. In fact, the results of various studies on distributed practice, variable practice, and retrieval practice suggest that learning and performance can be at odds, such that conditions that appear to degrade acquisition performance are often the very conditions that yield the most durable and flexible learning. Finally, research in metacognition suggests that fleeting gains during acquisition are likely to fool instructors and students into thinking that permanent learning has taken place, creating powerful illusions of competence.

That learning and performance are dissociable has widespread implications for theory, research, and practice. Any present (or future) comprehensive theory of learning and memory needs to distinguish, in some way, between the relatively permanent changes in behavior and knowledge that characterize long-term learning and transfer and the momentary changes in performance that occur during the acquisition of such behavior and knowledge. Likewise, researchers interested in elucidating factors that optimize learning should be cognizant of the possibility that the effects of manipulating a given variable might very well interact with retention interval—in other words, the variable might have differential influences on learning and performance. As such, we recommend that experimenters include both short- and long-term measures in their studies.

Finally, given that the goal of instruction and practice—whether in the classroom or on the field—should be to facilitate learning, instructors and students need to appreciate the distinction between learning and performance and understand that expediting acquisition performance today does not necessarily translate into the type of learning that will be evident tomorrow. On the contrary, conditions that slow or induce more errors during instruction often lead to better long-term learning outcomes, and thus instructors and students, however disinclined to do so, should consider abandoning the path of least resistance with respect to their own teaching and study strategies. After all, educational interventions should be based on evidence, not on historical use or intuition.

Acknowledgments

James S. McDonnell Foundation Grant 29192G supported the writing of this review. We thank Henry L. Roediger III for his insightful comments on an earlier version of this article.

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship or the publication of this article.

References

Abushanab, B., & Bishara, A. J. (2013). Memory and metacognition for piano melodies: Illusory advantages of fixed- over random-order practice. *Memory & Cognition*, *41*, 928–937.

Adams, J. A., & Reynolds, B. (1954). Effect of shift in distribution of practice conditions following interpolated rest. *Journal of Experimental Psychology*, *47*, 32–36.

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

Ariel, R., Hines, J. C., & Hertzog, C. (2014). Tests framing generates a stability bias for predictions of learning by causing people to discount their learning beliefs. *Journal of Memory and Language*, *75*, 181–198.

Armstrong, T. R. (1970, August). *Training for the production of memorized movement patterns* (Technical Report No. 26). Ann Arbor, MI: Human Performance Center of The University of Michigan.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*, 627–635.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321.

Baker, C. H. (1968). An evaluation of guidance in learning a motor skill. *Canadian Journal of Psychology*, *22*, 217–227.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging*, *4*, 3–9.

Battig, W. F. (1962). Paired-associate learning under simultaneous repetition and nonrepetition conditions. *Journal of Experimental Psychology*, *64*, 87–93.

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 215–244). New York, NY: Academic Press.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Voss (Eds.), *Topics in learning and performance* (pp. 131–159). New York, NY: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 23–44). Hillsdale, NJ: Erlbaum.

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.

Bendig, A. W. (1952). Latent learning in a water maze. *Journal of Experimental Psychology*, *43*, 134–137.

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.

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

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*, 392–402.

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, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

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

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & 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. (2011). On the symbiosis of learning, remembering, and forgetting. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 1–22). London, England: Psychology Press.

Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the mind* (pp. 242–244). Thousand Oaks, CA: Sage.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L.-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 109–133). Hove, England: Psychology Press.

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

Blodgett, H. C. (1929). The effect of the introduction of reward upon the maze performance of rats. *University of California Publications in Psychology*, *4*, 113–134.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*, 245–248.

Boutin, A., Badets, A., Salesse, R. N., Fries, U., Panzer, S., & Blandin, Y. (2012). Practice makes transfer of motor skills imperfect. *Psychological Research*, *76*, 611–625.

Boutin, A., Panzer, S., & Blandin, Y. (2013). Retrieval practice in motor learning. *Human Movement Science*, *32*, 1201–1213.

Buxton, C. E. (1940). *Latent learning and the goal gradient hypothesis*. Durham, NC: Duke University Press.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 484–496.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*, 279–283.

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. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*, 107–111.

Catalano, J. F., & Kleiner, B. M. (1984). Distant transfer in coincident timing as a function of variability of practice. *Perceptual & Motor Skills*, *58*, 851–856.

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.

Chasey, W. C., & Knowles, C. J. (1973). Effects of overlearning on retention and relearning of gross-motor skill by mentally retarded males. *Perceptual & Motor Skills*, *36*, 503–509.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23–56). Washington, DC: National Academies Press.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, *43*, 627–634.

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

Driskell, J. E., Willis, R. P., & Cooper, C. (1992). The effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615–622.

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.

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

Estes, W. K. (1955a). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369–377.

Estes, W. K. (1955b). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145–154.

Feijen, L., Hodges, N. J., & Beek, P. J. (2010). Acquiring a novel coordination skill without practicing the correct motor commands. *Journal of Motor Behavior*, *42*, 295–306.

Fitts, P. M. (1965). Factors in complex skill training. In R. Glasser (Ed.), *Training research and education* (pp. 177–197). New York, NY: Wiley.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *6*(Serial No. 40).

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Gilbert, T. F. (1957). Overlearning and the retention of meaningful prose. *Journal of General Psychology*, *56*, 281–289.

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 282–294.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.

Goode, M. K., Geraci, L., & Roediger, H. L. (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review*, *15*, 662–666.

Goode, S., & Magill, R. A. (1986). The contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport*, *57*, 308–314.

Green, P. D., Whitehead, J., & Sugden, D. A. (1995). Practice variability and transfer of a racket skill. *Perceptual & Motor Skills*, *81*, 1275–1281.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*, 505–513.

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, *36*, 212–224.

Guthrie, E. R. (1952). *The psychology of learning*. New York, NY: Harper & Row.

Hagman, J. D. (1983). Presentation- and test-trial effects on acquisition and retention of distance and location. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 334–345.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual & Motor Skills*, *78*, 835–841.

Hall, K. G., & Magill, R. A. (1995). Variability of practice and contextual interference in motor skill learning. *Journal of Motor Behavior*, *27*, 299–309.

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.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 290–296.

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.

Hodges, N. J., & Campagnaro, P. (2012). Physical guidance research: Assisting principles and supporting evidence. In N. J. Hodges & A. M. Williams (Eds.), *Skill acquisition in* *sport: Research, theory and practice* (2nd ed., pp. 150–169). London, England: Routledge.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*, 514–527.

Hull, C. L. (1943). *Principles of behavior*. New York, NY: Appleton-Century-Crofts.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension and competence. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–80). Washington, DC: National Academies Press.

Kane, J. H., & Anderson, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology*, *70*, 626–635.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.

Kantak, S. S., & Winstein, C. J. (2012). Learning-performance distinction and memory processes for motor skills: A focused review and perspective. *Behavioural Brain Research*, *228*, 219–231.

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

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

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*, 227–239.

Kelso, J. A. S., & Wallace, S. A. (1978). Conscious mechanisms in movement. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 79–116). New York, NY: Academic Press.

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual & Motor Skills*, *46*, 395–401.

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language*, *66*, 731–746.

Koriat, A. (1997). Monitoring one's knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

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

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

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, 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.

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 festschrift 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 learning. *Journal of Experimental Psychology: General*, *138*, 449–468.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in younger and older adults. *Psychology and Aging*, *25*, 498–503.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 989–998.

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

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology*, *12*, 71–78.

Krueger, W. C. F. (1930). Further studies in overlearning. *Journal of Experimental Psychology*, *13*, 152–163.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London, England: Academic Press.

Landin, D. K., Hebert, E. P., & Fairweather, M. (1993). The effects of variable practice on the performance of a basketball skill. *Research Quarterly for Exercise and Sport*, *64*, 232–237.

Lang, A. J., Craske, M. G., & Bjork, R. A. (1999). Implications of a new theory of disuse for the treatment of emotional disorders. *Clinical Psychology: Science and Practice*, *6*, 80–94.

Lee, T. D. (2012). Contextual interference: Generalization and limitations. In N. J. Hodges & A. M. Williams (Eds.), *Skill acquisition in sport: Research, theory, and practice II* (pp. 79–93). London, England: Routledge.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*, 277–287.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 730–746.

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (pp. 3–22). Amsterdam, The Netherlands: Elsevier.

Lee, T. D., Magill, R. A., & Weeks, D. J. (1985). Influence of practice schedule on testing schema theory predictions in adults. *Journal of Motor Behavior*, *17*, 283–299.

Lin, C. H., Wu, A. D., Udompholkul, P., & Knowlton, B. J. (2010). Contextual interference effects in sequence learning for younger and older adults. *Psychology and Aging*, *25*, 929–939.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 828–835.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction*, *4*, 91–115.

Martenuik, R. G. (1973). Retention characteristics of motor short-term memory cues. *Journal of Motor Behavior*, *5*, 249–259.

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, 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.

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

McCracken, H. D., & Stelmach, G. E. (1977). A test of the schema theory of discrete motor learning. *Journal of Motor Behavior*, *9*, 193–201.

McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York, NY: Psychology Press.

Melcher, R. T. (1934). Children's motor learning with and without vision. *Child Development*, *5*, 315–350.

Melnick, M. J. (1971). Effects of overlearning on the retention of a gross motor skill. *Research Quarterly*, *42*, 60–69.

Melnick, M. J., Lersten, K. C., & Lockhart, A. S. (1972). Retention of fast and slow learners following overlearning of a gross motor skill. *Journal of Motor Behavior*, *4*, 187–193.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.

Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.

Merbah, S., & Meulemans, T. (2011). Learning a motor skill: Effects of blocked versus random practice: A review. *Psychologica Belgica*, *51*, 15–48.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102–116.

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., & 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.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133.

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology*, *66*, 206–209.

Pigott, R. E., & Shapiro, D. C. (1984). Motor schema: The structure of the variability session. *Research Quarterly for Exercise and Sport*, *55*, 41–45.

Postman, L. (1962, February 23). Retention as a function of degree of overlearning. *Science*, *135*, 666–667.

Postman, L., & Tuma, A. H. (1954). Latent learning in human subject. *American Journal of Psychology*, *67*, 119–123.

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*, 644–667.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*, 70–80.

Reynolds, B. (1945). A repetition of the Blodgett experiment on "latent learning." *Journal of Experimental Psychology*, *35*, 506–516.

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.

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

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

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*, 1323–1330.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498.

Roller, C. A., Cohen, H. S., Kimball, K. T., & Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor plasticity. *Cognitive Brain Research*, *12*, 341–352.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology*, *67*, 605–610.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*, 225–260.

Schmidt, R. A. (2003). Motor schema theory after 27 years: Reflections and implications for a new theory. *Research Quarterly for Exercise and Sport*, *74*, 366–375.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Schmidt, R. A., & Lee, T. D. (2011). *Motor control and learning: A behavioral analysis* (5th ed.). Champaign, IL: Human Kinetics.

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.

Seward, J. P. (1949). An experimental analysis of latent learning. *Journal of Experimental Psychology*, *39*, 177–186.

Shea, C. H., & Kohl, R. M. (1990). Specificity and variability of practice. *Research Quarterly for Exercise and Sport*, *61*, 169–177.

Shea, C. H., & Kohl, R. M. (1991). Composition of practice: Influence on the retention of motor skills. *Research Quarterly for Exercise and Sport*, *62*, 187–195.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179–187.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 907–912.

Skinner, B. F. (1938). *The behavior of organisms*. New York, NY: Appleton-Century-Crofts.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior*, *22*, 153–163.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*, 405–412.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1582–1593.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*, 341–358.

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. (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.

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (in press). Metamemory and education. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory*. Oxford University Press.

Spence, K. W., & Lippitt, R. (1946). An experimental test of the sign-gestalt theory of trial and error learning. *Journal of Experimental Psychology*, *36*, 491–502.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.

Stelmach, G. E. (1969). Efficiency of motor learning as a function of intertrial rest. *Research Quarterly*, *40*, 198–202.

Stelmach, G. E., Kelso, J. A. S., & Wallace, S. A. (1975). Preselection in short-term motor memory. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 745–755.

Ste-Marie, D. M., Clark, S. E., Findlay, L. C., & Latimer, A. E. (2004). High levels of contextual interference enhance handwriting acquisition. *Journal of Motor Behavior*, *36*, 115–126.

Stevenson, H. W. (1954). Latent learning in children. *Journal of Experimental Psychology*, *47*, 17–21.

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.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837–848.

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

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.

Thorndike, E. L. (1914). *The psychology of learning*. New York, NY: Teachers College Press.

Tolman, E. C. (1932). *Purposive behavior in animals and men*. New York, NY: Century.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 198–208.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology*, *4*, 257–275.

Tsutsui, S., & Imanaka, K. (2003). Effect of manual guidance on acquiring a new bimanual coordination pattern. *Research Quarterly for Exercise and Sport*, *74*, 104–109.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*, 429–442.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381–391.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

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.

Van Merrienboer, J. J. G., de Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual & Motor Skills*, *84*, 784–786.

Wallis, R., Elliot, B., & Koh, M. (2002). The effect of a fast bowling harness in cricket: An intervention study. *Journal of Sports Sciences*, *20*, 495–506.

Waters, R. H. (1930). The influence of large amounts of manual guidance upon human maze learning. *Journal of Genetic Psychology*, *4*, 213–228.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571–580.

Winstein, C. J., Pohl, P. S., & Lewthwaite, R. (1994). Effects of physical guidance and knowledge of results on motor learning: Support for the guidance hypothesis. *Research Quarterly for Exercise and Sport*, *70*, 316–323.

Wrisberg, C. A., & Mead, B. J. (1983). Developing coincident timing skill in children: A comparison of training methods. *Research Quarterly for Exercise and Sport*, *54*, 67–74.

Wulf, G. (1991). The effect of type of practice on motor learning in children. *Applied Cognitive Psychology*, *5*, 123–134.

Wulf, G., & Schmidt, R. A. (1988). Variability in practice: Facilitation in retention and transfer through schema formation or context effects? *Journal of Motor Behavior*, *20*, 133–149.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*, 185–211.

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, *42*, 1373–1383.

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.

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