

# **Neuroscience in Education**

The good, the bad and  
the ugly

Edited by

Sergio Della Sala

and

Mike Anderson

**OXFORD**  
UNIVERSITY PRESS

## Chapter 8

# **Applications of cognitive science to education**

Henry L. Roediger, III, Bridgid Finn and  
Yana Weinstein

### **Overview**

We present five topics (sets of findings) from cognitive psychology that can be directly applied to the classroom and/or to students' study strategies outside of class. These include: (1) retrieval practice through testing, which enhances learning more than restudying does; (2) spaced periods of study of the same topic relative to massed study of that topic; (3) interleaving of different topics of study rather than blocked studying of the same topic; (4) improved metacognitive monitoring for students (that is, teaching students to assess accurately the state of their own knowledge); and (5) teaching in ways that will facilitate transfer of learning to novel situations. Current educational practice often uses the less beneficial of these strategies (e.g. repeated massed study of the same topic), and thus principles derived from cognitive psychology can help to encourage better practices in the classroom and in individual study periods outside of class. Although these five topics are only some of the ways cognitive psychologists have shown that student learning can be improved, we believe they represent techniques that are most applicable to improving education today. These bodies of knowledge have stood the test of time (some of the topics have been studied for over 100 years) and their validity is not in doubt. The trick for educators and students is to find ways to apply these strategies in beneficial ways.

The goal of this volume is to provide insights from neuroscience and cognitive science aimed at improving educational practice. Our chapter falls squarely into the latter camp, that of cognitive science (although we do point to some neuroscientific findings in places). We extrapolate research from purely behavioural cognitive psychology and make recommendations for educational practice. In a few cases, our suggestions have already been implemented in the classroom, at least on an experimental basis. Of course, neural processes underlie any cognitive processes, so in a broad (and vacuous) sense our chapter can be conceived as being about neuroscience and education. However, we certainly do not make this claim. We seek to generalize findings from behavioural studies to possible educational practice.

## 8.1 Cognition and its relation to neuroscience and education

Because the orientation of our chapter differs from some other chapters in the book, we pause to consider and defend our perspective. In 1997, John Bruer wrote an article entitled 'Education and neuroscience: A bridge too far'. He reviewed the enthusiasm sweeping educational circles in the mid 1990s for applying neuroscience to education but argued that it was misplaced, often based on simplistic interpretations or even misunderstandings of the claims of neuroscience. He wrote: 'Educational applications of brain science may come eventually, but as of now neuroscience has little to offer teachers in terms of informing classroom practice. There is, however, an applied science of mind, cognitive science, which can serve as a basic science of the development of an applied science of learning and instruction' (p. 4). Bruer argued that the link between neuroscience and education represented a 'bridge too far', and he suggested that two shorter spans must be constructed before researchers could traverse the divide between neuroscience and education. The first bridge that must be constructed is from cognition to instruction: What are the cognitive underpinnings of instructional practice and how can these be made more effective? Once a good start has been made at this task, the second bridge must link from cognition to neural circuitry: what neural circuits underlie the cognitive processes that are linked to instructional practice? In the past 15 years, good progress has been made on this latter front in some areas (e.g. brain mechanisms for reading and for straightforward mathematical computation), but certainly much more work remains to be done. Our chapter aims mostly at building the first of Bruer's bridges, the link between cognition and instruction.

Other chapters in this volume build a strong case for neuroscience and education, or what Carew and Magsamen (2010) have called neuroeducation. They defined it as 'a nascent discipline that seeks to blend the collective fields of neuroscience, psychology, cognitive science and education to create a better understanding of how we learn and how this information can be used to create more effective teaching methods, curricula and educational policy' (p. 685). By this very broad definition, our chapter does fit under the umbrella of neuroeducation.

## 8.2 Organisation of the chapter

We have selected five topics from cognitive psychology of learning and memory (see further discussion in Chapter 7, this volume) that we believe have special import for education: retrieval practice (testing) of information as a mnemonic enhancer; spacing of study episodes to improve performance relative to massed (back-to-back) study episodes; interleaving of various topics of study rather than blocking them; metacognition (the knowledge of one's own cognitive processes and how they affect learning); and transfer of knowledge (applying learned information in a new domain). These topics have been thoroughly studied in laboratory settings (some more than others, of course), and we believe all five are ripe for application in classroom settings.

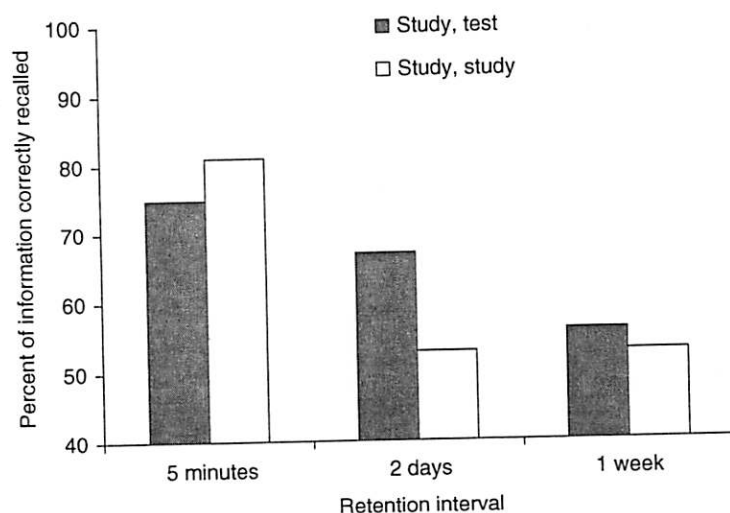
## 8.3 Retrieval practice via testing

Educators generally give tests to assess students' learning and to assign grades. In addition, administrators give standardized tests to assess student learning on a common measurement scale (say, within a state or country or even across countries). Controversies abound in the use of testing, especially for this latter purpose. In this section, we argue that tests can serve another important purpose, too—they can foster better learning and retention as well as assessing it. The act of taking a test can greatly enhance knowledge of the tested material. Experimental psychologists have known this phenomenon, called the testing effect, for over 100 years; Abbott (1909) seems to have been the first to note it through experimentation, although thinkers back to Aristotle have

supposed that the principle is true. In his essay on 'Memory and reminiscence', Aristotle wrote: 'Exercise in repeatedly retrieving a thing strengthens the memory'. It only took another 2500 years or so for his claim to receive empirical confirmation. However, Aristotle did not get everything right. He thought the heart to be the repository of memories, and the phrase of 'learning by heart' provides a vestige of this claim.

A straightforward experiment by Roediger and Karpicke (2006) shows the power of testing with only a single test. They had students read fact-filled passages about a variety of topics (the sun, sea otters). Shortly after reading a passage, students either took an initial test on the passage (lasting 7 minutes) or they read it again for 7 minutes. For the test, they were given the title of the passage (sea otters) and asked to recall as much as they could. The passage was divided into idea units (fundamental ideas or concepts in each sentence) so that it could be reliably scored. On this first test, students in the relevant condition recalled 70% of the idea units. Of course, in the condition that involved restudying the whole passage, students were exposed to 100% of the units. This re-exposure condition would seem to be at an advantage over the testing condition because students did not perfectly recall the passage during the test.

The critical part of the experiment occurred later, when students were given a final criterial test. This test was given to different groups of subjects either shortly after the manipulation (about 5 minutes later), after 2 days, or after 7 days. The results are shown in Figure 8.1. At the short retention interval, the repeated study condition produced greater recall than the study/test condition. This outcome shows what students have long known—cramming (repeated study) just before a test elevates performance on that test. However, performance on the two delayed tests showed just the opposite, with the study/test condition outperforming the study/study condition by a wide margin. This outcome represents the testing effect: taking a test generally produces better performance on a later test than does repeated study. Notice that the increased delay caused greater forgetting, but the forgetting was much smaller for the condition that took an initial test than for the condition in which students repeatedly studied the information. Testing seems to insulate against forgetting (see, too, Carpenter, Pashler, Wixted & Vul, 2008; Wheeler, Ewers &



**Fig. 8.1** Data from Roediger and Karpicke (2006a, experiment 1). On the test taken 5 minutes after study, material that students repeatedly studied was recalled better than material that was studied once and then tested. On the two delayed tests, however, the pattern reversed: studying and taking a test led to better performance than did repeated studying.

Buonanno, 2003). One idea to explain the benefit of testing is that it permits students to practice the skill they will need later, a form of transfer appropriate processing (e.g. Roediger, Gallo & Geraci 2002). More specifically, research by Zaromb and Roediger (2010) supports the idea that in practising retrieval of large sets of information with few cues, students organise the information more effectively (create good retrieval plans) than they do when they repeatedly study the information, and this retrieval plan aids later organised recall.

Another fact to glean from Figure 8.1 is the nature of the interaction between the study/study and study/test conditions and retention interval, viz., a crossover interaction. As noted previously, the interaction arises (descriptively) because repeated study leads to better initial performance, but the initial test slows forgetting, thus leading to the crossover interaction. Forgetting does occur in the study/test condition, of course, but it is more gradual. In this experiment a testing effect occurred only on a delayed test, but in many other experiments testing effects are found in the same experimental session (e.g. Carrier & Pashler, 1992 among many others).

One boundary condition for the testing effect is the level of initial performance on the test. In the Roediger and Karpicke (2006) experiment just described, students recalled 70% of the idea units on the first test. But what if the first test were delayed 2 weeks and performance had dropped to, say, 30%? Then the rereading condition would have been exposed to 100% of the idea units 2 weeks later and the testing group to only 30%. Would the testing effect still occur? The answer is no, or not usually. Initial level of performance is one of the boundary conditions, as Kang, McDermott and Roediger (2007) showed. However, the testing effect does emerge even with low initial test performance if feedback is given. That is, even if students are tested at a delay and recall only 30% of the material, if they receive feedback a testing effect will still occur. This fact may seem odd; after all, giving feedback simply re-exposes students to the material as in the comparison rereading condition. However, the requirement for students to expend effort in trying to retrieve seems to exert a positive effect when they receive feedback. Izawa (1970) referred to this as the potentiating effect of testing, arguing that taking a test enhances the information encoded and stored from the next study opportunity relative to a pure restudy condition (see Arnold & McDermott, in preparation). This test-potentiating effect can even occur when a test is given before any material is studied (Kornell, Hays & Bjork, 2009; Richland, Kornell & Kao, 2009).

In all the cases just cited, there seems to be something about making an effort to retrieve that facilitates performance, and Pyc and Rawson (2009) have produced evidence consistent with the idea that expending retrieval effort may underlie the testing effect in certain situations. Nonetheless, the important point for practical purposes is that feedback should be given whenever possible after tests. Feedback corrects errors students make (Pashler, Cepeda, Wixted & Rohrer, 2005) and can benefit even correct answers if they are ones made with low confidence (Butler, Karpicke & Roediger, 2008). The usual advice is that feedback should be given as soon as possible after a test, preferably immediately. However, in experiments directly comparing immediate to (somewhat) delayed feedback, delayed feedback produces larger benefits on later tests (Butler, Karpicke & Roediger, 2007). The reason may be that delayed feedback permits spaced presentation of material and, as we shall see in the next section of the chapter, spaced practice benefits retention relative to massed practice. Providing immediate feedback amounts to massed presentation and thus produces less long-term benefit.

A large amount of research on the benefits of testing has been published in recent years (see Roediger & Butler, 2011 and Roediger, Agarwal, Kang & Marsh, 2010 for short and long reviews, respectively). The important point for present purposes is that testing provides a robust benefit in later recall relative to restudying. Further, testing effects have been obtained in simulated classroom settings (e.g. Butler & Roediger, 2007). The power of testing effects has led researchers to bring testing into actual classrooms to see if their use can improve student achievement

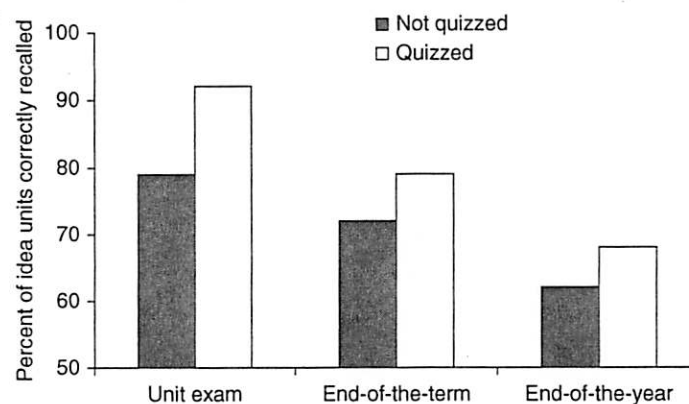
(e.g. Carpenter, Pashler & Cepeda, 2009; McDaniel, Agarwal, Huelser, McDermott & Roediger, 2011; Roediger, Agarwal, McDaniel & McDermott, 2011).

We will use one experiment from Roediger et al. (2011) to illustrate the possible effectiveness of retrieval practice via testing as a tool for improving learning in a middle school classroom. The research was fully integrated into a social studies class for 11–12-year-olds during the first term of an academic year. Topics covered included major civilizations (e.g. ancient Egypt, India, China and so on). The experiments were conducted within-students, with some facts about each topic randomly assigned for quizzing and another set of facts assigned to the no-quiz condition. The teacher had six classes, and there was a different random assignment of materials to condition for each class; the teacher was unaware of which facts were assigned to which condition (she was outside the classroom when the materials were quizzed). A research assistant performed the quizzing using a student response (clicker) system that permitted immediate feedback (but also permitted the assistant to obtain scores). The quizzes and final criterial tests were all multiple choice, because that is the evaluation mode the teacher preferred.

The quizzing was a common part of students' daily routine. From their point of view, they read (or not) the assigned material, the teacher covered it in class and they reviewed (or not) the material before the chapter exam. The quizzes given in class were viewed as practice, and students seemed to enjoy them (they often complained on the days when quizzes were not given—'why can't we use the clickers today?'). Facts that were tested were quizzed three times, once before the teacher presented the material in class, once the day after she presented it and once just before the chapter exam.

The scores of interest occurred during three later tests: a chapter exam, an end-of-the-term exam and a final exam at the end of the school year in May (approximately 7–8 months after the manipulation). In the case of chapter and term exams, the grades counted for the student grades; that is, our dependent measures were collected as part of the exams the students were taking for grades. At the end of the year this feature was not included; the students received a surprise test on material covered many months earlier.

The results are shown in Figure 8.2, where it can be seen that a testing effect occurred on all three tests. On the chapter exams, items that had been quizzed were correctly answered 91% of the time relative to 74% of the time for non-quizzed items. The teacher reported that her classes



**Fig. 8.2** Performance of middle school students at three different time-points on science material that was previously quizzed in class, or was not (from McDaniel, Agarwal, Huelser, McDermott & Roediger, 2011). Material that was previously quizzed produced consistently better performance on all three tests.

usually scored in the mid 70s on her tests, so the fact that we could raise performance from this level to the low 90s means that (on the grading scale used in class) students rose from a C+ to an A– average. Obviously, on the delayed tests given at the end of the semester and the end of the year, performance dropped (forgetting occurred), but the testing advantage remained. Even at the end of the school year items quizzed in October and November were better remembered than items not quizzed. We suspect that if we had been able to use production (recall tests) such as short-answer tests, we might have found even larger effects due to the greater retrieval effort they engender (see Kang et al., 2007).

In sum, retrieval practice via testing leads material to be better retained than does either no practice or repeated study of material. This finding has important implications for study strategies of individual students (e.g. Karpicke, 2009) and for classroom practice (see, too, McDaniel, Agarwal, Huelser, McDermott & Roediger, 2011; McDaniel, Thomas, Agarwal, McDermott & Roediger, in preparation). In addition, material learned via retrieval practice also seems to transfer to novel situations (when the material is tested in other ways) and so represents more than ‘teaching to the test’ (Butler, 2010). Retrieval practice serves as a critical mnemonic booster. We turn now to other techniques to improve learning and memory.

## 8.4 Spaced practice

The spacing effect is the robust finding that distributing practice, by spacing out several study episodes over time rather than massing them all at once, can substantially boost long-term learning. Ebbinghaus (1885/1964) noted that, ‘with any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time’ (p. 89). Since this early observation, the benefits of spaced practice on long-term retention have been established across a wide range of populations and domains. Besides hundreds of positive findings with college students, spacing effects have been shown in studies with children (Son, 2010; Toppino, 1991), older adults (Balota, Duchek & Paullin, 1989), non-human animals (Davis, 1970) and even amnesic patients (Cermak, Varfallie, Lanzoni, Mather & Chase, 1996). The advantages of spacing have also been shown to occur across a vast range of domains and tasks, including motor learning, classical and operant conditioning, implicit measures of memory, recognition memory, paired associate learning, free recall, text processing, statistics learning and vocabulary acquisition (see, e.g. Cepeda, Vul, Pashler, Wixted & Rohrer, 2006; Dempster, 1988, 1996; Glenberg, 1979; Greene, 1990 and Hintzman, 1974, provided reviews). Finally, the spacing effect occurs whether the presentation modality is visual or auditory (Melton, 1970), and even when the two presentations occur in distinct modalities (Hintzman, Block & Summers, 1973).

In general, the benefits of spacing grow with practice and increase with the length of the lag between presentations (Pavlik & Anderson, 2005). In free recall experiments, the number of items that occur between presentations shows a systematic relationship with later recall, with greater lags leading to better recall (e.g. Glenberg, 1976; Madigan, 1969; Melton, 1967). Melton presented subjects with words two times. Some of the words were massed—in other words, presented twice in a row. Other repetitions were spaced by 2, 4, 6, 20, or 40 different interpolated words. Results showed that the probability of recall on a later test steadily increased with the lag between presentations. It is important to note, however, that there may be some differences in the temporal parameters of the effect among different populations and experimental conditions (Hintzman, 1974; Toppino & DeMesquita, 1984; Wilson, 1976).

Spacing of learning has been shown to be an effective technique in improving the cognitive function of patients with various types of neurological dysfunction (Schacter, Rich & Stamp, 1985).

Spacing provides an advantage relative to other techniques in that it does not require effort on the part of the patients. Judged against other techniques that have been shown to aid learning in patient groups, such as imagery, organization, or verbal labelling (e.g. Cermak, 1975, Cermak, Reale & DeLuca, 1977; Gianutsos & Gianutsos, 1979), spacing is relatively undemanding on the patient's cognitive resources (Bjork, 1979; Schacter et al., 1985). After exposure to spaced training, some patients show spontaneous use of spaced retrieval (Schacter et al., 1985), demonstrating that spaced training can be implemented and retained by patient populations.

In the classroom, spaced practice has long been regarded as a way to enhance learning. In William James's (1899/1958, pp. 93–94) essays to teachers and students he advises,

Cramming seeks to stamp things in by intense application immediately before the ordeal. But a thing thus learned can form but few associations. On the other hand, the same thing recurring on different days, in different contexts, read, recited on, referred to again and again, related to other things and reviewed, gets well wrought into the mental structure... There is no moral turpitude in cramming. It would be the best, because the most economical, mode of study if it led to the results desired. But it does not...

Some of the earliest studies of distributed practice were done in educational settings. Pyle (1913) drilled 8–9-year-olds on math problems once a day for 10 days, or once in the morning and once in the afternoon for 5 days. When practice was extended over more days, learning was more effective. Similarly, Smith and Rothkopf (1984) found that distributing a lesson over 4 days was more effective than presenting it for the same amount of total time on one day.

The spacing effect has been demonstrated with a variety of educationally relevant materials. Rea and Modigliani (1985) showed better spelling and mathematics performance following spaced practice. Vocabulary learning is facilitated by spaced relative to massed practice (Bahrick & Phelps, 1987; Dempster, 1987), with the gains shown to be quite long lasting. For example, Bahrick and Phelps (1987) tested people's retention of Spanish–English vocabulary words that they had learned 8 years earlier. They found that people who had initially learned the vocabulary pairs in spaced training sessions showed much better retention than the group who had received massed training.

In a recent study, Kornell and Bjork (2008) tested the effectiveness of massed versus spaced practice on students' ability to learn concepts and categories. Successive presentations of category exemplars might result in superior learning by allowing people to notice similarities amongst the exemplars, whereas targeting within-category similarities might be more difficult if exemplars are spaced out. Participants studied six paintings by each of 12 artists, such as Braque and Seurat. Six of the artists' paintings were studied in massed presentation, and six of the artists' paintings were studied with spaced presentations. On the final test, participants were shown new paintings from each of the 12 artists and were required to indicate whom they thought was the artist for each painting. Classification of the new paintings to artists was superior following spaced practice. Interestingly, despite the large advantage for the spaced items, when students were asked to judge which type of presentation they thought was most effective, they reported that massed study was more effective than spaced study. We will return to this point shortly.

Spaced practice is equally, if not more effective when the second presentation is a test trial or a trial involving a test plus a restudy opportunity for feedback (Glover, 1989; Landauer & Bjork, 1978; Whitten & Bjork, 1977). So, for instance, students study an item (say, a face–name pair) once, and on the second presentation the item is tested (e.g. the face is given and subjects try to produce the name) before the whole pair is presented again. As we have discussed in the previous section, testing provides a considerable boost to long-term retention. It is possible to further enhance the testing effect by spacing test trials. Glover (1989) examined the extent to which

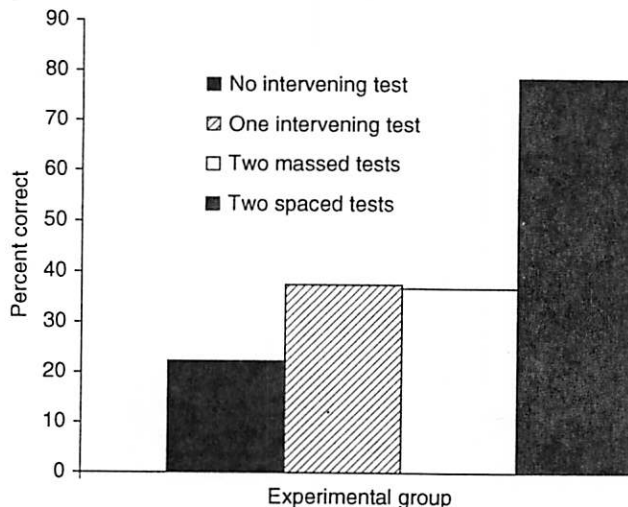


taking massed and spaced tests improved performance on a later test. Seventh graders studied materials that were part of their normal science course. Before the final test, the students received intervening tests according to one of four conditions: a single test, two massed tests, two spaced tests, or no intervening test. As can be seen in Figure 8.3, reconstructed from the data reported in Glover (1989), final test performance was worst in the no intervening test condition and best in the two spaced tests condition. Importantly, performance did not differ between the single test and two massed tests conditions. Additional tests were more effective than a single test only when they were spaced apart, possibly because of the additional processing involved in the second spaced test relative to the second massed test.

In another demonstration of the benefits of spaced testing, Landauer and Bjork (1978) asked participants to learn face-name pairs. The pairs were repeated according to different rehearsal (retrieval) patterns. The initial presentation was the face and a first and last name. On subsequent presentations, the pairs received a test-type practice in which the face was shown with the first name and a blank for participants to attempt recall of the last name. These presentations were either massed or spaced. In contrast to the massed condition, the items that had received spaced test-type practice showed much better recall on a final test (see Roediger & Karpicke, 2011 for a review).

Over 20 years ago, Dempster (1988) described the irregular use of the spacing effect in education as 'a failure to apply the results of psychological research'. In spite of the many demonstrations of enhanced learning of educationally relevant materials as a result of spacing, students are reluctant to space their learning. Why? For one reason, massed practice may just feel better. With massed study, processing each repetition is fluent whereas processing during a spaced practice can feel more challenging. Similarly, when tests are massed, recall attempts are easier and so success rates are high.

Metacognitive assessments about whether something will be remembered in the future are often based on feelings of fluency, even though it can be a misleading index of later retention



**Fig. 8.3** Percent correct in a flower labelling task as a function of the number of intervening tests (0, 1 or 2) and the distribution of those tests (massed or spaced). One test was better than 0 tests, and a second test was only more beneficial than a single test if the two tests were spaced rather than massed. Reported in Glover (1989; Experiment 3).

(Benjamin, Bjork & Schwartz, 1998). Because learning with massed practice feels faster and easier than when practice is spaced (even when massing produces inferior performance), people rate massing as the more effective learning strategy, as they did in the Kornell and Bjork (2008) study cited earlier (see, too, Baddeley & Longman, 1978 and Zechmeister & Shaughnessy, 1980). This confidence is misplaced, because often the easier something is to process, the worse recall will be later (Bjork, 1994; cf. Koriatic, 2008). Many studies have shown that when given the opportunity to control their own learning, people will usually choose to mass rather than space their study (e.g. Landauer & Ross, 1977), and this is especially so when learning feels more difficult to the subjects (Toppino, Cohen, Davis & Moors, 2009; but see Benjamin & Bird, 2006). People will space items that they judge to be well learned (Son, 2004, Toppino et al., 2009), although young children tend to mass their practice (Son, 2005).

In sum, most studies show that spaced practice provides benefits relative to massed practice, yet too often teachers provide information in massed fashion in class (one topic, then another topic, then a third and so on) rather than trying to space out reviews of previously covered topics. In addition, students tend to study the same way, concentrating or massing their efforts on one topic after another. As Dempster (1988) noted, spacing of information would doubtless be a boon to education if it could be more widely adopted in schools and if students could be taught to study using this principle.

## 8.5 Interleaving of topics

Just as spacing is more beneficial to learning than massing, interleaving different sets of materials during study can promote learning as compared to blocking materials together. The interleaving of materials always involves their spacing, but interleaving goes beyond spacing by having students study one skill, or type of problem, or set of material, and then cycle through the various skills, etc., in various orders. (Spacing can be achieved in other ways besides interleaving, as discussed later.) This is opposed to practising the same skill or set of material repeatedly and then moving on to another. For example, if children are learning the skills of multiplication and division, one strategy for teaching (and a common one) is to have them do 25 multiplication problems and then 25 division problems. This blocked presentation often promotes accurate and efficient performance after a short period of practice. The interleaving alternative is to randomly alternate multiplication and division problems. Under these conditions, accuracy and speed grow more slowly over trials, but the benefits are more long lasting. That is, in most studies retention is much better following interleaved practice than following massed practice. However, as with spacing, students and teachers generally prefer massed training because students seem to learn so quickly under these conditions.

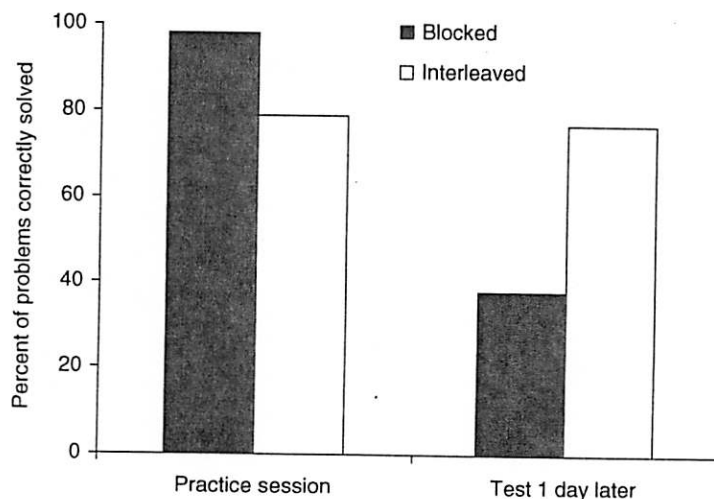
The benefits of interleaving have been demonstrated with physics problems (Rohrer & Taylor, 2007), mathematics problems (Le Blanc and Simon, 2008) and even learning to identify artists (Kornell & Bjork, 2008—as described in the previous section). As an example, Rohrer and Taylor had students learn how to find the volume of four geometric solids. Students then practised solving problems for each solid, in one of two conditions: either the problems for one solid were all solved before moving on to the next solid or the problems for all four solids were randomly intermixed. On a follow-up test, students in the intermixed conditions performed over three times better than students who had practised the problems blocked by each type of solid. Despite these benefits, an overwhelming majority of mathematics (and other) textbooks block their practice problems.

Before any strong conclusions can be drawn, however, it is important to distinguish between the benefit arising from spacing described in detail previously, and the benefit derived from interleaving

specifically, because (as just noted) interleaving by definition involves spacing. Taylor and Rohrer (2010) designed a study to do just that by creating blocked tasks that involved spacing and comparing performance on interleaved problems (with spacing). The basic question they addressed is 'Does interleaving problems benefit their later retention more than spacing the problems?'

Rohrer and Taylor (2010) had students aged 9–10 years solve four different types of mathematics problems. In the task, students were shown a picture of a prism and told how many base sides it had. They then had to use one of four formulas to calculate either the number of faces (number of base sides + 2), the number of corners (number of base sides  $\times$  2), the number of edges (number of base sides  $\times$  3), or the number of angles (number of base sides  $\times$  6). In the learning phase, students either solved blocks of the same problem (e.g. they repeatedly calculated the number of corners in different prisms before moving on to calculate the number of angles in a different set of prisms), or they practised all four types of problems in an interleaved fashion. This is the typical comparison of blocking and interleaving, but in their experiment there was a twist imposed on the blocked condition because it included a 30-second filler task (puzzles unrelated to the mathematics problems of interest) between sets of the same type of problem. The reason for inserting these filler tasks was to match the gap between problems of the same kind with that of the interleaved group. In other words, because there were four different problem types, if a student in the interleaved condition attempted a problem that involved calculating the number of faces first, they would attempt the problems involving corners, edges and angles before being given a further problem where they had to calculate the number of faces. At 10 seconds per problem, the gap between the first and second problems involving calculating the number of faces was 30 seconds—exactly the same length of time that students in the blocked group performed a filler task. Thus the blocked condition was changed to be the same as a typical spaced condition in a spacing effect experiment; therefore, any improved learning in the interleaved condition can be attributed to interleaving specifically as opposed to spacing more generally.

Figure 8.4 shows average performance during learning in the two conditions as well as performance on a delayed test the next day. Despite the fact that the blocked condition involved spacing



**Fig. 8.4** Performance of students aged 9–10 years on four different types of mathematical problems during a practice session and on a test a day later, from an experiment by Taylor and Rohrer (2010). In the blocked condition, each type of problem was practised separately. In the interleaved condition, all four problem types were randomly intermixed during practice.

in this experiment, interleaving had its usual effects on performance. The bars on the left of Figure 8.4 show that interleaving slowed learning relative to blocking, as it usually does. That is partly why both students and teachers do not like to interleave practice—it slows initial learning. Yet performance on a test a day after learning was far better in the interleaved compared with the blocked condition, as shown by the bars on the right side of Figure 8.4. Comparing the two sides of Figure 8.4, it is apparent that there was great forgetting in the blocked condition and none in the interleaved condition. Thus, interleaving has a benefit on learning (as measured on delayed tests) over and above spacing. Presumably there is something about the effort of switching between tasks and reinstating the mental procedures necessary for solving them that creates better retention for the long term in the interleaved condition.

Although interleaving has not been studied from a neuroscientific perspective (yet), a similar concept has been studied extensively under a different name. Multitasking—defined by Delbridge (2001) as ‘engaging in frequent switches between individual tasks’—has been studied extensively by neuroscientists, with the focus largely on identifying and locating the costs of switching between tasks. This line of research paints a negative picture of interleaving, because performance and brain activity are measured during the learning phase and never during later recall, when the benefits of interleaving appear. Neurological theories of why interleaving works would be purely speculative at this stage, but it is reasonable to propose that the extra cognitive load caused by switching between tasks or types of problems in an interleaved study phase actually benefits retrieval.

The take away point for the classroom is that the customary practice of blocking should be diminished in favour of more interleaved and spaced practice. Students learning to print letters should not necessarily do A 25 times, then B and so on. Massed practice creates quick learning but also produces quicker forgetting relative to interleaved (or spaced) practice.

## 8.6 Metacognition

Surveys of university students’ study practices show that they differ greatly from those that cognitive psychologists recommend (Karpicke, Butler & Roediger, 2009; Kornell & Bjork, 2007; McCabe, 2011). Why might this be so? One fact might be that many good learning strategies proposed by cognitive psychologists tend to hurt performance in the short term while improving it in the long term. Students may be more focused on short-term processes—ones that they can readily judge—and so be guided by their immediate intuitions and beliefs. Central to this paradox is the issue of whether students are able to monitor and regulate their own learning in an effective manner.

The primary model of metacognition (Nelson & Narens, 1990) proposes that people evaluate the progress of their learning and use this evaluation to regulate study strategies. Students use their evaluations to plan and control their activities such as allocating additional study time to material that seems to be learned less well. Research has shown that metacognitive monitoring plays a key role in how people control their study (e.g. Metcalfe & Finn, 2008; Thiede, 1999; but see Weinstein, Finn, McDermott & Roediger, in preparation). The efficacy of this process of course hinges on the accuracy of the metacognitive assessments as well as the integrity of the strategies that are implemented.

In the sections above we touched on how students’ metacognitive evaluations can be insensitive to the beneficial effects that a number of learning strategies (retrieval practice, spacing and interleaving) have on learning. In all three cases, the ‘easy’ version of the task that seems most natural—repeated massed study—leads to poorer long-term retention than the alternative strategy (testing, spaced or interleaved study). In some cases, metacognitive inaccuracies may occur due to reliance on heuristics that would be predictive in other instances, as Kahneman (2003) has argued.

For example, massed study works fine for an immediate test so students use it; however, massed practice is poor for performance on delayed tests, but students may not realize that fact and use the strategy even when it is inappropriate.

Metacognitive inaccuracies can also arise from reliance on cues like familiarity or fluency that may not be diagnostic of whether or not something is learned. For example, rereading may lead to fluent processing of the material and this fluency can lead to overconfidence that one knows the material that seems so fluent and familiar (e.g. Benjamin et al., 1998; Glenberg, Sanocki, Epstein & Morris, 1987; Metcalfe, Schwartz & Joaquim, 1993; Reder & Ritter, 1992). Thus, overconfident students who reread the material repeatedly may stop studying before they have truly mastered material because of faulty monitoring (e.g. Metcalfe & Finn, 2008; Zechmeister & Shaughnessy, 1980). Because monitoring is tightly linked to the control of study, students need to learn how to evaluate their current state of knowledge by using cues or tests that are diagnostic of their state of learning; otherwise they can fall prey to illusions of knowing. For example, if students test themselves at a reasonable delay from learning, they can discover whether or not they can produce the studied material when it is needed. If they do not quiz themselves but simply continue to restudy, the material may seem fluent and easy, but they may not be able to retrieve it after a delay (e.g. Roediger & Karpicke, 2006).

One technique used to study the development of metacognitive knowledge during learning is called the judgement of learning (JOL) paradigm. The JOL paradigm simply asks students to predict whether they will later be able to recall or recognize a fact that has been learned. If you learn today that Jefferson City is the capital of Missouri, you might be asked to judge how well you could recall this fact in a month. If people are asked to perform JOLs while studying material, the JOLs are usually not very accurate (though usually not at chance, either); however, if the JOLs are made after a short delay when attempting to retrieve the correct target information, they are much more predictive of later success (Nelson & Dunlosky, 1994). During study, while students are examining the material, the facts being learned may seem easy and fluent; yet when students judge learning after even a brief delay, they may appreciate how difficult the material is to retrieve and thus not be fooled. As noted earlier, metacognitive accuracy can also be improved following a test (Finn & Metcalfe, 2007; 2008; Koriat & Bjork, 2006) by asking students to think of reasons why their answer might be wrong (Koriat, Lichtenstein & Fischhoff, 1980), by asking students to summarize what they have learned before making their judgements (Thiede & Anderson, 2003), and by warning them about the possibility of bias (Jacoby & Whitehouse, 1989).

When metacognitive accuracy is good, control of study efforts is effective (Nelson, 1992). When people control their own study time, their performance is better than when they are given a random amount of time (Atkinson, 1972; Mazzone & Cornoldi, 1993) or when they are given the opposite value of the amount of time they chose for study (Kornell & Metcalfe, 2006). Nelson (1993) had students study vocabulary pairs and then make JOLs. After making the JOLs, students were given additional study trials for some items. One group of participants received additional study time for the items that they had given the lowest JOLs, which indicated that they were sure that they would not remember them on a later test. Another group restudied the items that they had given the highest JOLs, items that they were sure they would remember. The control group restudied items that were made up of half high JOL and half low JOL items. The group that had restudied the low JOL items performed the best on the follow-up test, indicating that people were sensitive to what they did and did not know and that studying the items that they did not know benefited learning.

A large body of literature has shown that people often choose to spend the most time studying the items that they have given the lowest JOLs (see Son & Metcalfe, 2000 for a review). This pattern fits with the idea that people will study until there is no longer a discrepancy between their judgement

of learning and the learning objective. This idea is called the discrepancy reduction model of study time allocation and was proposed by Dunlosky and Thiede (1998) to explain why students tend to study material they judge as more difficult for longer periods of time. However, a discrepancy reduction strategy does not always work. For some (by definition, difficult) material, students can be given a large amount of time to study and still not master the material. Nonetheless, when permitted to control their study strategies, students persist in processing this difficult material even when studying has diminishingly small effects, a process Nelson and Leonesio (1988) called *labouring in vain*. People often do not know when to stop studying (or perhaps to change strategies and study in a different way).

When study time is limited, as it usually is, students may not always study the most difficult material. Work by Metcalfe and her collaborators has shown that people do not always choose to study the most difficult items when faced with learning a list of pairs that vary in difficulty (Metcalfe, 2002; Metcalfe & Kornell, 2003, 2005; Kornell & Metcalfe, 2006). Under time pressure, study strategies shift. With time constraints, items that are of moderate difficulty are prioritized over items that are already known or items that are too difficult. These moderately difficult items are in what Metcalfe and her collaborators have called the 'region of proximal learning.' That is, students will go for the low-hanging fruit—the material that has not been learned but that seems to be learnable in a short amount of time. In support of this model, Metcalfe and Kornell (2005) showed that when given the option, students prioritize easier unknown items in preference to extremely difficult items.

Additional support for this region of proximal learning idea comes from research by Kornell and Metcalfe (2006). In one series of experiments, students were given an initial test on Spanish–English vocabulary pairs. From the set of incorrect items participants selected half for restudy. Results showed that performance on the final test was better when people's study choices were honoured and they were given the items that they had selected for restudy relative to when their choices were dishonoured and they were given the items that they had not selected. Importantly, the items that they chose were the easiest items (according to norms) but ones as yet not learned.

Some neuroscientific evidence does bear on the issue of metacognition when learning. Converging evidence from patient and neuroimaging studies suggest that a number of regions in the prefrontal cortex are activated in metacognitive monitoring, and that the areas involved in metamemory do not completely overlap with those areas involved in remembering (see, e.g. Pannu & Kaszniak, 2005; Schwartz & Bacon, 2008; and Shimamura [2008] for a review). Using functional magnetic resonance imaging (fMRI), Kao, Davis and Gabrieli (2005) investigated the neural circuitry that mediates JOLs. Participants studied pictures and predicted how well they would later remember each scene. Results showed that activation in the medial temporal lobe was associated with actual, but not predicted, recall success. Activation in the left ventromedial prefrontal cortex was associated with predictive accuracy, but not with actual performance. Areas in the lateral and dorsomedial prefrontal cortex were associated with both successful recall and JOLs. The results point to at least somewhat distinct processes involved in remembering and judgements about remembering.

Research with patients with damage to their frontal lobes provides further support for the idea that the prefrontal cortex is important for metacognitive monitoring and control. Lesions in the prefrontal cortex have been associated, in particular, with impairments in feeling of knowing (FOK) judgements (Janowsky, Shimamura & Squire, 1989; Pannu & Kaszniak, 2005; Schnyer et al., 2004). During a FOK task participants make a judgement about the likelihood that they will be able to recognize the answer to the question that they were not able to correctly answer at that time. An example is: 'What is the capital of Kentucky?'. If the answer is unknown, a FOK judgement is made. Later the participant is asked to recognize the answer from among several



alternatives: Lexington, Louisville, Frankfort or Paducah? The FOK judgement taps into people's access to partial information about the target, which might be comprised of the first letter (Koriat & Leiblich, 1975), syntactic characteristics (Miozzo & Caramazza, 1997) or other kinds of knowledge (Metcalf & Finn, 2011). Frontal patients have much worse accuracy in their predictions of future performance (e.g. Janowsky et al., 1989; Pinon, Allain, Kefi, Dubas & LeGall, 2005; Vilkki, Servo & Surma-aho, 1998).

The frontal lobes develop throughout childhood and are thought to mature in adolescence or even early adulthood (e.g. Welsh & Pennington, 1988; Welsh, Pennington & Groisser, 1991; and see Romine & Reynolds, 2005 for a review) and metacognitive capacities follow a similar developmental trajectory (e.g. Schneider, 2010). For example, while young children's metamemory judgements can be quite inaccurate, by the time they reach middle school their ability to monitor memory appropriately seems to be in place (e.g. Metcalf & Finn, submitted; Schneider & Lockl, 2002). For learning to proceed successfully though, the output of metacognitive evaluations needs to be implemented into good study strategies. Research suggests that the ability to monitor accurately may appear before the student knows how to use that information effectively (Metcalf & Finn, submitted; Schneider, 2010). Metacognitive control can lag behind accurate monitoring, but it does show age-related improvements (e.g. Dufresne & Kobasigawa, 1989; Lockl & Schneider, 2004). While adults and older children spend more time studying items that they do not know, younger children make less adaptive choices and may need assistance to choose optimally (Metcalf & Finn, submitted).

Because metacognitive processes are used to regulate children's learning behaviours, it is crucial that their monitoring is accurate and their control decisions are adaptive. Several studies have shown that overconfidence is related to poor exam performance (Bol & Hacker, 2001; Kruger & Dunning, 1999). Moreover, overconfidence in a particular study strategy may encourage students to persist in using ineffective strategies (Hacker, Bol & Bahbahani, 2008). Low-achieving students face a double burden: Over a wide spectrum of academic domains, those students with the lowest levels of performance show the greatest overconfidence (e.g. Kruger & Dunning, 1999; and see Hacker et al., 2008 for a review). Fortunately, metacognitive strategies can be successfully trained (e.g. Brown & Campione, 1990), and research points toward the importance of classroom practices in teaching students how to regulate their own learning (see Schneider & Pressley, 1997 for a review). Specifically, instructing students as to why a particular study strategy is useful increases use of that strategy (O'Sullivan & Pressley, 1984). Pressley and collaborators have shown that the most effective teachers are those who incorporate instruction about how to use metacognition to select and modify study strategies (e.g. Pressley, Goodchild, Fleet & Zajchowski, 1989).

As Hacker and others (2008) have noted, however, the large bulk of research on metacognition has been conducted in laboratory settings with hopes that it can be generalized to more naturalistic educational contexts. The needed translational research is just now underway. Findings from the lab do not always transfer neatly into the classroom. In classroom settings, teachers may need to provide more information on how to use metacognition appropriately, especially with children and with low-achieving students. By explicitly training metacognitive monitoring and providing feedback about strategy selection, educators can improve how students use metacognition to learn (Hacker, 2004; Nietfeld, Cao & Osborne, 2005).

## 8.7 Transfer of learning

The techniques described so far in this chapter—testing, spacing, interleaving—will not be of much use unless the knowledge and techniques acquired during their implementation can be transferred to other materials and situations. Do these techniques lead to encapsulated knowledge

or do they lead to learning that is flexible and can be transferred to new contexts? *Transfer* has been a buzzword in educational policy ever since Edward Thorndike and Robert Woodworth began studying the topic over 100 years ago (Thorndike & Woodworth, 1901a, 1901b, 1901c). Today many educational organizations would agree that 'A main reason for formal education is to facilitate learning in situations outside school' (Klausmeier, 1961, p. 352). The exact definition of transfer has been long debated and continues to be discussed (see, e.g. Barnett & Ceci, 2002; Beach, 1994). In the context of education, transfer can refer to the production of the same piece of information in response to two differently worded questions, but also to the application of problem-solving skills acquired in one context to another. The former sort of transfer is often called near transfer and the latter is termed far transfer, with the distance terms referring to how similar the new task is to one that was trained. If transfer is achieved when the new task is only a slight modification of the original one, then this is a case of near transfer. When the two tasks are conceptually and procedurally different and yet transfer is achieved, it is said to be far transfer. Not surprisingly, many studies find near transfer to be quite robust, but relatively few have reported far transfer (although such studies do exist, as we shall see). In light of the effective study techniques described earlier, transfer could also refer to the propensity to use these techniques with new materials once their efficiency has been experienced.

Despite educators' general agreement that transfer of learning is crucial, surprisingly little evidence of such transfer exists in the lab, much less in the schools. In part, this difficulty arises from differences in definitions of transfer (and in near and far transfer). In Thorndike and Woodworth's (1901) original work on the topic entitled 'The influence of improvement in one mental function upon the efficiency of other functions', the authors were disappointed to find zero evidence of transfer. For example, subjects trained to estimate the areas of rectangles showed no transfer to estimating the areas of other shapes. This led the authors to conclude that 'Improvement in any single mental function rarely brings about equal involvement in any other function, no matter how similar . . .' (p. 247). More recently, this conclusion was echoed by Detterman (1993), who branded transfer an epiphenomenon and argued strongly that no real evidence of transfer had been produced by the literature. On the other side of the debate, Schwartz, Bransford and Sears (2005) argued that the classic definitions of far transfer are too narrow and lead to incorrect conclusions of failed transfer in many studies. According to their view, failure of transfer occurs because researchers are expecting their students to directly map a learned procedure onto an entirely new context rather than modifying learned procedures to suit the new situation. Schwartz et al. further suggested that a more important facet of transfer involves improvement in the ability to learn new ways of solving problems. With this revised definition, transfer can be identified in many situations, and the picture looks less bleak.

Nonetheless, the distinction between near and far transfer, referring to the degree to which the initial learning and transfer contexts differ, has allowed important advances in research on transfer. Educators are most concerned with far transfer (i.e. transfer to a context largely dissimilar to the initial learning context), but even the definition of far transfer has been problematic. For instance, Hamers, de Koning and Sijtsma (1998) claimed to have shown far transfer when the 8–9-year-olds trained on attribute classification transferred that skill to an intelligence quotient (IQ) test. However, Barnett and Ceci (2002) argued that since both the trained task and the transfer task were from the same *domain* (paper and pencil tests) the authors were really demonstrating near transfer. In an attempt to prevent what they considered to be misrepresentation of far transfer, Barnett and Ceci proposed a taxonomy of transfer, a framework for evaluating past experiments and designing new ones. They identified two major factors—content and context—with the latter consisting of a set of six dimensions that could be used to determine whether a given



situation involves near or far transfer: *knowledge domain*, *physical context*, *temporal context*, *functional context*, *social context* and *modality*. For instance, the *functional context* dimension describes the motivational mind set involved in an activity. Answering a geography question in the context of a school quiz and then remembering the same fact during a conversation with friends would thus involve far transfer due to the different functions that the piece of information served in the two cases. Figure 8.5 provides examples of near and far transfer involving each of the six dimensions.

Barnett and Ceci (2002) did describe two studies that demonstrated far transfer by their criteria. Fong, Krantz and Nisbett (1986) showed that statistical training in college courses transferred to an unexpected phone survey that required use of statistical knowledge. Likewise, Ceci and Ruiz (1993) found that real-life racetrack handicapping skills transferred to a laboratory-based stock market experiment; better handicapping skills led to better performance in the laboratory experiment that required use of the skills in a totally different context. However, most dimensions outlined by Barnett and Ceci (e.g. *social context*, or learning in collaboration with others at school but then coming up with the information alone during a test), have not received as much attention in the literature. Whether such transfer can be demonstrated must await future research.

Working closely with the Barnett and Ceci (2002) framework, Butler (2010) designed a series of experiments to vary only the knowledge domain dimension. After studying passages on topics such as 'tropical cyclones' and 'vaccines', students answered inferential questions within the same knowledge domain (for near transfer) or a different domain (for far transfer). As an example of near transfer, after answering a question on the general uses of vaccines, students would attempt a question about the purpose of a particular vaccine. The far transfer question could be about the function of a vehicle, and students were expected to draw upon their understanding of the human body gained from the vaccine passage to correctly explain the processes occurring inside the

	Near ← → Far				
<b>Knowledge domain</b>	Mouse vs. rat	Biology vs. botany	Biology vs. economics	Science vs. history	Science vs. art
<b>Physical context</b>	Same room at school	Different room at school	School vs. research lab	School vs. home	School vs. the beach
<b>Temporal context</b>	Same session	Next day	Weeks later	Months later	Years later
<b>Functional context</b>	Both clearly academic	Both academic but one nonevaluative	Academic vs. filling in tax forms	Academic vs. informal questionnaire	Academic vs. at play
<b>Social context</b>	Both individual	Individual vs. pair	Individual vs. small group	Individual vs. large group	Individual vs. society
<b>Modality</b>	Both written, same format	Both written, multiple choice vs. essay	Book learning vs. oral exam	Lecture vs. wine tasting	Lecture vs. wood carving

**Fig. 8.5** Examples of near and far transfer for the six domains defined by Barnett and Ceci (2002; adapted from their figure 1B).

vehicle—clearly a very different domain to the one they had studied! Crucially, the design also involved a comparison between testing and restudying, to investigate which technique led to greater transfer (either near or far transfer). That is, students either repeatedly studied facts or they studied the facts and took tests on them; then they took a final test that assessed near or far transfer as described above.

Butler (2010) found that initial testing led to better transfer than repeated studying (a testing effect), and further that the testing effect occurred on both tests of near transfer and far transfer. This is an important finding for the testing effect, since it demonstrates that testing is not only beneficial for retaining the specific facts that are rehearsed at retrieval, but also improves transfer of knowledge to new domains.

In another study, Rohrer, Taylor and Sholar (2010) had 9–11-year-olds study or practise retrieving city locations. Afterwards, the students were given a task in which they had to determine which city they would drive through on a given route. Performance on the driving task was better after retrieval practice than after repeated study of the city locations, another indication of transfer. Evidence of improved transfer has also been demonstrated for other techniques described above such as spacing; that is, spaced study leads to greater transfer on a later test than does massed study (Helsdingen, van Gog & van Merriënboer, 2009).

Atherton (2007) proposed a neuroscientific theory of transfer. According to this theory, transfer will occur insofar as the brain regions activated in different contexts are interconnected. Atherton cites demonstrations of transfer from music ability to language processing (e.g. Anvari, Trainor, Woodside & Levy, 2002) along with findings of overlap between brain regions involved in music and language processing (e.g. Koelsch et al., 2004) as indirect empirical evidence of this hypothesis. From an alternative perspective, Haskell (2001) proposed that the human brain evolved to support transfer. Clearly, these new theories need to be tested empirically before they can guide educational practice; for now, Barnett and Ceci's (2002) taxonomy of transfer and carefully specified behavioural work are leading the way.

Although studies of transfer of learning in educational contexts lag behind other realms of research reviewed in this chapter, this is a critical topic for future research. Most education aims to be useful, to provide students with skills and knowledge they will need throughout life. Thus educators will need to know how to teach (and have students study) to create flexible knowledge that can be retained and used over long periods of time.

## 8.8 Conclusion

This chapter has shown how five different topics studied by cognitive psychology can have implications for education: retrieval practice through testing; spaced periods of study of the same topic; interleaving of different domains of study; improving metacognitive monitoring of students; and teaching in ways that will facilitate transfer of learning to novel situations. These five topics are some of those that cognitive psychology can contribute to educators to help guide practice. However, we believe that these are only some of the tools that are needed. The strategies outlined here assume students will learn readily and will be motivated to learn, but sadly these assumptions are often not met. Indeed, among full-time university students only about 11% of first years spend as much time preparing for class as their professors expect (NSSE, 2008). Developmental and social psychologists, among others, must be enlisted to help understand how students can become motivated to learn and want to apply the strategies described here. The task for us all is long, and we can hope that neuroscientific approaches will increasingly shed light on these issues in the future.

## References

- Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs*, 11, 159–77.
- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing and early reading ability in preschool children. *Journal of Experimental Child Psychology*, 83, 111–30.
- Arnold, K. M., & McDermott, K. B. (in preparation). Test potentiated learning. Manuscript in preparation.
- Atherton, M. (2007). *A Proposed Theory of the Neurological Limitations of Cognitive Transfer*. Paper presented at the annual meeting of the American Educational Research Association, 9–13 April, Chicago, IL.
- Atkinson, R. C. (1972). Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology*, 96, 124–29.
- 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, 627–35.
- Bahrack, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 344–49.
- Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age related differences in the spacing of repetitions and retention interval. *Psychology & Aging*, 4, 3–9.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological-Bulletin*, 128(4), 612–37.
- Beach, K. D. (1994). *A Sociohistorical Alternative to Economic and Cognitive Transfer Metaphors for Understanding the Transition from School to Work*. Paper presented at the annual meeting of the American Educational Research Association, April, New Orleans, LA.
- Benjamin, A. S., & Bird, R. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory & Language*, 55, 126–37.
- 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.
- Bjork, R. A. (1979). An information-processing analysis of college teaching. *Educational Psychologist*, 14, 15–23.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bol, L., & Hacker, D. (2001). The effect of practice tests on students' calibration and performance. *Journal of Experimental Education*, 69, 133–51.
- Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking or a context by any other name. In D. Khun (Ed.), *Contributions to Human Development*, 21, 108–25.
- Bruer, J. (1997). Education and the brain: A bridge too far. *Educational Researcher*, 26(8), 4–16.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118–33.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273–81.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 918–28.
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514–27.
- Carew, T. J., & Magsamen, S. (2010). Neuroscience and education: An ideal partnership for producing evidence based solutions to guide 21st century learning. *Neuron*, 67, 685–88.

- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology*, 23, 760-71.
- Carpenter, S. K., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36, 438-48.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 632-42.
- Ceci, S. J., & Ruiz, A. (1993). The role of context in everyday cognition. In M. Rabinowitz (Ed.), *Applied Cognition* (pp. 164-83). Hillsdale, NJ: Erlbaum.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354-80.
- Cermak, L. S. (1975). Imagery as an aid to retrieval for Korsakoff's patients. *Cortex*, 77, 163-69.
- Cermak, L. S., Reale, L., & DeLuca, D. (1977). Korsakoff patient's nonverbal vs verbal memory: Effects of interference and mediation on rate of information loss. *Neuropsychologia*, 15, 303-310.
- Cermak, L. S., Verfaellie, M., Lanzoni, S., Mather, M., & Chase, K. A. (1996). The spacing effect on the recall and recognition ability of amnesic patients. *Neuropsychology*, 10, 219-27.
- Davis, M. (1970). Effects of interstimulus interval length and variability on startle-response habituation in the rat. *Journal of Comparative and Physiological Psychology*, 78, 260-67.
- Delbridge, K. (2001). *Individual Differences in Multi-tasking Ability: Exploring a Nomological Network*. Unpublished Doctoral Dissertation, Michigan State University, East Lansing.
- Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79, 162-70.
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, 43, 627-34.
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of Perception and Cognition*, Vol. 10: *Memory*, (pp. 317-44). San Diego, CA: Academic Press.
- Detterman, D. K. (1993). The case for prosecution: Transfer as an epiphenomenon. In D. K. Detterman, & R. J. Sternberg, *Transfer on Trial: Intelligence, Cognition, and Instruction*, (pp. 1-24). Norwood, NJ: Ablex Publishing Corporation.
- Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47, 274-96.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, 98, 37-56.
- Ebbinghaus, H. (1885/1964). *Memory: A Contribution to Experimental Psychology*. Oxford: Dover.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 238-44.
- Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, 58, 19-34.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, 18, 253-92.
- Gianutsos, R., & Gianutsos, J. (1979). Rehabilitating the verbal recall of brain-injured patients by mnemonic training: An experimental demonstration using single-case methodology. *Journal of Clinical Neuropsychology*, 2, 117-35.
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1-15.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95-112.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119-36.

- Glover, J. A. (1989). The 'testing' phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392-99.
- Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1004-1011.
- Hacker, D. J. (2004). Self-regulated comprehension during normal reading. In R. B. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 775-79). Newark, DE: International Reading Association.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration in classroom contexts: The effects of incentives, reflection, and attributional style. *Metacognition and Learning*, 3, 101-21.
- Hamers, J. H. M., De Koning, E., & Sijtsma, K. (1998). Inductive reasoning in the third grade: Intervention promises and constraints. *Contemporary Educational Psychology*, 23, 132-48.
- Haskell, R. E. (2001). *Transfer of Learning: Cognition, Instruction and Reasoning*. New York: Academic Press.
- Heldsingen, A. S., van Gog, T., & van Merriënboer, J. J. G. (2009). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, 21(1), 126-36.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior*, 12, 229-38.
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, 83, 340-44.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126-35.
- James, W. (1899/1958). *Talks to Teachers on Psychology: And to Students on Some of Life's Ideals*. New York: Holt.
- Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, 27, 1043-56.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58, 697-720.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology*, 19, 528-58.
- Kao, Y. C., Davis, E. S., & Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience*, 8(12), 1776-783.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469-86.
- 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, 471-79.
- Klausmeier, H. J. (1961). *Educational Psychology: Learning and Human Abilities*. New York: Harper.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302-307.
- Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, 36, 416-28.
- Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, 34, 959-72.
- Koriat, A., & Lieblich, I. (1975). Examination of the letter serial position effect in the 'TOT' and the 'Don't Know' states. *Bulletin of the Psychonomic Society*, 6, 539-41.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, 14, 219-24.

- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the 'enemy of induction'? *Psychological Science*, 19, 585–92.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35, 989–98.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 32, 609–22.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–34.
- Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Harris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 625–32). New York: Academic Press.
- Landauer, T. K., & Ross, B. H. (1977). Can simple instruction to space practice improve ability to remember a fact? An experimental test using telephone numbers. *Bulletin of the Psychonomic Society*, 10, 215–218.
- Le Blanc, K., & Simon, D. (2008). *Mixed Practice Enhances Retention and Jol Accuracy for Mathematical Skills*. Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.
- Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology*, 1, 153–69.
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, 8, 828–35.
- Mazzoni, G., & Cornoldi, C. (1993). Strategies in study-time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, 122, 47–60.
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39, 462–76.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399–414.
- McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (in preparation). Quizzing promotes transfer of target principles in middle school science: Benefits on summative exams. Manuscript in preparation.
- Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, 158, 532.
- 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.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131, 349–63.
- Metcalfe, J., & Finn, B. (submitted). Metacognition and control of study choice in children. Manuscript submitted for publication.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review*, 15, 174–79.
- Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, 132, 530–42.
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463–77.
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 851–61.
- Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology: Learning Memory and Cognition*, 23, 1410–23.

- National Survey of Student Engagement (NSSE) (2008). *Promoting engagement for all students: The imperative to look within (2008 results)*. Bloomington: Indiana University Center for Postsecondary Research.
- Nelson, T. O. (1992). *Metacognition: Core Readings*. Needham Heights, MA: Allyn & Bacon.
- Nelson, T. O. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, 122, 269–73.
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2, 325–35.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the 'labor-in-vain effect'. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 47–86.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–41.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the classroom. *Journal of Experimental Education*, 74, 7–28.
- O'Sullivan, J., & Pressley, M. (1984). The completeness of instruction and strategy transfer. *Journal of Experimental Child Psychology*, 38, 275–88.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: a review. *Neuropsychology Review*, 15, 105–30.
- Pashler, H., Cepeda, N., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 3–8.
- Pavlik, P. I. Jr., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29, 559–86.
- Pinon, K., Allain, P., Kefi, M. Z., Dubas, F., & Le Gall, D. (2005). Monitoring processes and metamemory experience in patients with dysexecutive syndrome. *Brain Cognition*, 57, 185–88.
- Pressley, M., Goodchild, F., Fleet, J., & Zajchowski, R. (1989). The challenges of classroom strategy instruction. *Elementary School Journal*, 89(3), 301–42.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60, 437–47.
- Pyle, W. H. (1913). Economical learning. *Journal of Educational Psychology*, 4, 148–58.
- Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 4, 11–18.
- Reder, L. M., & Ritter, F. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 435–51.
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, 15, 243–57.
- Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New Frontiers in Applied Memory* (pp. 13–49). Brighton: Psychology Press.
- Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 103, 299–414.
- 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. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–55.
- Roediger, H. L. & Karpicke, J. D. (2011). Intricacies of spaced retrieval: A resolution. In A.S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: Essays in Honor of Robert A. Bjork*. New York: Psychology Press.

- Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels of processing framework. *Memory*, 10, 319–32.
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems improves learning. *Instructional Science*, 35, 481–98.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 233–39.
- Romine, C. B. & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: findings from a meta-analysis. *Applied Neuropsychology*, 12, 190–201.
- Schacter, D. L., Rich, S. A., & Stampp, M. S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced retrieval technique. *Journal of Clinical and Experimental Neuropsychology*, 7, 79–96.
- Schneider, W. (2010). Metacognition and memory development in childhood and adolescence. In H. Slaatas Waters & W. Schneider (Eds.), *Metacognition, Strategy Use, and Instruction* (pp. 54–81). Guilford Press: New York.
- Schneider, W., & Lockl, K. (2002). The development of metacognitive knowledge in children and adolescents. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied Metacognition* (pp. 224–57). Cambridge: Cambridge University Press.
- Schneider, W., & Pressley, M. (1997). *Memory Development Between 2 and 20*. Hillsdale, NJ: Erlbaum.
- Schnyer, D. M., Verfaellie, M., Alexander, M. P., Lafleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling of knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, 42, 957–66.
- Schwartz, B. L., & Bacon, E. (2008). Metacognitive neuroscience. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of Memory and Metamemory: Essays in Honor of Thomas O. Nelson*, (pp. 355–71). Psychology Press: New York.
- Schwartz, D. L., Bransford, & J. D., Sears, D. L. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective* (pp. 1–51). Greenwich, CT: Information Age Publishing.
- Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory*, (pp. 373–90). New York: Psychology Press.
- Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, 1, 341–58.
- Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 601–604.
- Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies. *Journal of General Psychology*, 132, 347–63.
- Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 255–62.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 204–21.
- Taylor, K., & Rohrer, D. (2010). The effects of interleaving practice. *Applied Cognitive Psychology*, 24, 837–48.
- Thiede, K. W. (1999). The importance of accurate monitoring and effective self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6, 662–67.
- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–60.
- Thorndike, E. L., & Woodworth, R. S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247–61.
- Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions: II. The estimation of magnitudes. *Psychological Review*, 8, 384–95.



- Thorndike, E. L., & Woodworth, R. S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation, and discrimination. *Psychological Review*, 8, 553–64.
- Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition*, 19, 159–67.
- Toppino, T. C., Cohen, M. S., Davis, M., & Moors, A. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1352–58.
- Toppino, T. C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology*, 37, 637–48.
- Varma, S., McCandliss, B. D., & Schwartz, D. L. (2008). Scientific and pragmatic challenges for bridging education and neuroscience. *Educational Researcher*, 37(3), 140–52.
- Vilkkii, J., Servo, A., & Surma-aho, O. (1998). Word list learning and prediction of recall after frontal lobe lesions. *Neuropsychology*, 12(2), 268–77.
- Weinstein, Y., Finn, B., Roediger, H. L., & McDermott, K. B. (in preparation). The effect of retention interval expectancy on study time and accuracy. Manuscript in preparation.
- Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, 4(3), 199–230.
- Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window of prefrontal function in children. *Developmental Neuropsychology*, 7(2), 131–49.
- Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11, 571–80.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: The effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16, 465–78.
- Wilson, W. P. (1976). Developmental changes in the lag effect: An encoding hypothesis for repeated word recall. *Journal of Experimental Child Psychology*, 22, 113–22.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38, 995–1008.
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 15, 41–44.