Zulkiply, N. & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: moderation by the difficulty of category discriminations. *Memory and Cognition, 41*(1), 16–27. https://doi.org/10.3758/s13421-012–0238-9

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction, 22*(3), 215–221. https://doi.org/10.1016/j.learninstruc.2011.11.002

17 Correcting Student Errors and Misconceptions

Elizabeth J. Marsh and Emmaline Drew Eliseev

Introduction

Overview and History

Mistakes are common in education and sometimes are even encouraged as part of the learning process. Young children, for example, are often instructed to "invent" spellings while writing, as this approach leads them to write longer texts (albeit with more spelling errors; Clarke, 1988). Many textbook chapters prompt students to answer "prequestions" about material that has not yet been covered, with the goal of guiding their reading (Pressley et al., 1990). Learning to debug errors in computer code is considered integral to learning to program (e.g., Klahr & Carver, 1988). In general, the prevailing zeitgeist is for educational practice to encourage learners to have a growth mindset and to accept mistakes as part of the learning process (Yeager & Dweck, 2012).

Such practices differ greatly from what was considered ideal practice prior to the 1970s. Perhaps most notably, the behaviorist B. F. Skinner argued strongly against a "trial-and-error" view of learning in favor of avoiding errors when possible (Skinner, 1953). The emphasis was on *errorless learning* (Terrace, 1963), meaning that learning conditions should allow students to learn without ever making a mistake (under the logic that errors had the potential to interfere with learning the correct responses). This perspective is an "early selection" model of errors, in that it aims for errors to never occur in the first place. Such an approach is insufficient for two reasons: First, in at least some instances, making a mistake may help learning (e.g., Kornell, Hays, & Bjork, 2009). Second, and more practically, it is virtually impossible to prevent a learner from ever making mistakes, necessitating a mechanism for error correction (a "late correction" model).

The Importance of Error Correction

Understanding how to correct errors is crucial, as uncorrected errors have the potential to be problematic. For example, students who do not receive feedback on their answers on a multiple-choice test are at risk for later reproducing some of the multiple-choice lures they selected (for a review, see Marsh et al., 2007). A student who incorrectly selects

"mice" as the answer to the multiple-choice question "With the increase in nutrias in the U.S., which animal's population decreased?: (a) beavers, (b) mice, (c) moles, (d) muskrats" will be more likely to answer "mice" to a later short-answer version of the question ("With the increase in nutrias in the U.S., which animal's population decreased?") than if they had never answered the multiple-choice question at all. This problem occurs with college students, high school students (Fazio, Agarwal et al., 2010), and even elementary school children (Marsh, Fazio, & Goswick, 2012). The problem is not question-specific, in that incorrectly selecting "gravitation" in response to the question "What biological term describes fish slowly adjusting to water temperature in a new tank?: (a) acclimation, (b) gravitation, (c) maturation, (d) migration" affects responses to conceptually similar questions, such as "Animals that thicken their fur during winter are exhibiting what biological phenomenon?" (Marsh et al., 2007). Fortunately, there is a relatively simple solution: Tell students the correct answer (Butler & Roediger, 2008) – previewing one of the main messages of this chapter.

Uncorrected errors have consequences beyond believing a single falsehood is true; conceptual misunderstandings can interfere with new learning of other related information. Students enter the classroom with naïve theories about how things work, based on their own experiences as well as portrayals in film and television, and such beliefs are thought to impede acquisition of fundamental concepts in fields such as chemistry (e.g., Nakhleh, 1992), physics (e.g., Carey, 1986), and computer programming (e.g., Clancy, 2004). It is for this reason that teachers often try to anchor new concepts on familiar objects and systems (e.g., using the solar system to teach about the atom), to make sure they are using the correct mental model for the problem.

Defining Feedback

Broadly, we consider feedback to be any information that has the potential to affirm or update a student's knowledge (as opposed to feedback regarding student behavior or motivation, for example). Our definition of feedback is not tied to a particular type of error and it can take many forms, including but not limited to marks indicating whether an answer is right or wrong, the correct answer to a problem or question, an explanation of why an answer is the correct one, or a new way of thinking about a complex system. It can occur at different grain sizes, providing only the needed correction or contextualizing the correction within a larger unit (e.g., an entire textbook chapter or lesson). Feedback could come from another person (such as a teacher), from places such as the Internet or a book, or even from oneself (such as when one looks up the answer to check their work). In short, we endorse a broad definition of feedback in this chapter, with the different instantiations tied together by the *need to correct something* – meaning that even entire lessons may serve as feedback if the student enters with a misconception.

Methods for Studying Error Correction

Overview

There are hundreds of demonstrations of the power of feedback. Feedback is typically considered to be one of the most powerful tools in the teacher's toolbox, yielding a large effect size (average Cohen's *d* greater than 0.7; Hattie, 2012, 2015). The benefits of providing feedback are larger than those of other common recommendations to teachers, such as having students spread out learning over time (spacing; $d = 0.60$), create concept maps ($d = 0.64$), review worked examples ($d = 0.37$), and engage in peer tutoring ($d = 0.55$) (Hattie, 2015).

One of our favorite experimental examples comes from Pashler and colleagues (2005) because they separated the effects of feedback on maintaining correct answers versus correcting errors. That is, they argued that some of the inconsistent effects in the literature likely occurred because these two types of responses were treated as the same, when feedback should have less of an impact on correct answers (given that, by definition, there is nothing to correct). Their participants received two chances to study a list of twenty Luganda–English word pairs (e.g., *leero–today*), before taking a translation test (*leero–?*). On this initial test, the groups performed similarly (as expected), translating about 40 percent of words correctly – meaning that there were plenty of errors to be corrected. Critically, some subjects received no feedback about their answers, others were told whether each translation was correct or not, and a third group was told the correct translation of each Luganda word (additional controls were run to rule out possible confounds but we will not discuss those further). Of interest was the learner's ability to maintain their correct responses, as well as to correct their errors, both relatively immediately and after a delay.

While participants did forget some of their correct translations over the course of one week, this effect was similar across feedback conditions. Error correction, in contrast, depended on the feedback condition: Errors were only corrected after receiving answer feedback, not correct/incorrect feedback, and this pattern held both immediately and on a test one week later (although the expected forgetting happened over the course of a week). These effects were very large – receiving correct answer feedback following a mistake improved final performance by 494 percent. Such powerful effects are consistent with past reviews of the literature showing large benefits of feedback.

Pashler and colleagues' results highlight our first piece of advice: When students make an error, it is better to provide the correct answer than to simply mark it as incorrect. We reached a similar conclusion when we tested students on what they learned from reading short passages describing history, geography, and science (Fazio, Huelser et al., 2010). Answer feedback is even the better choice when correcting errors made on a multiple-choice test (Marsh et al., 2012). This finding with multiple-choice tests is particularly discouraging about the usefulness of right/wrong feedback, as such feedback does provide information in this case – it allows the learner to winnow down the remaining choices. But of course this benefit is

contingent on learners' ability to correctly eliminate lures, leading to the finding shown in Figure 17.1: Following errors on a multiple-choice test, receiving right/wrong feedback led to intermediate performance on a final test, between the levels observed following no feedback versus answer feedback. The benefit of right/wrong feedback dropped as narrowing down the choices became harder (with three remaining choices, for example, as opposed to one) – after being told one made an error on a four-alternative multiple-choice question, performance was not much higher than if one had received no feedback at all.

In another study, we examined whether receiving right/wrong feedback (and the knowledge of what one does vs. does not know) indirectly benefits learning, by guiding learners' future study efforts – but, unfortunately, our data did not support this idea. After reading texts and receiving feedback, readers received a second chance to read the passages, at their own pace. However, all readers benefited from rereading, even if they had not received any feedback at all. One possibility is that right/wrong feedback often does not provide any additional information to the learner – there are many cases where learners have a good sense of what they do versus do not know without feedback, such as translations of foreign vocabulary or simple facts (e.g., Hart, 1967). However, there are also cases where people do not realize they are making mistakes, something we turn to in the next section of this chapter.

Errors Made with Confidence

We know intuitively that not all errors are the same – it is one thing to correct an incorrect translation and another to correct a misunderstanding of why seasons occur. These two example errors differ in many ways, including but not limited to the
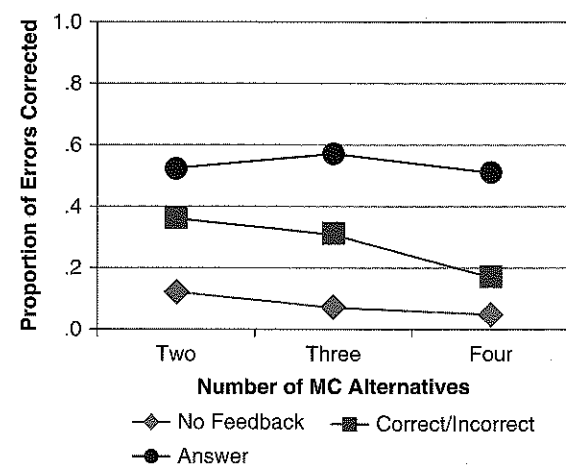
Figure 17.1 *Proportion of errors corrected on the final test based on feedback condition (no feedback, correct/incorrect, answer feedback) and the number of multiple-choice alternative answer choices (two, three, four) on an initial multiple-choice test (after Marsh et al., 2012)*

complexity of the error (a simple factual error vs. a conceptual misunderstanding of a system), confidence in one's response (which likely will be lower for the failed translation than one's faulty explanation of the seasons), content domain (language vs. science), and so on.

As a starting point, we can examine other errors that are similar to the incorrect translations in most ways but differ in one key aspect. For example, many people believe that George Washington had wooden teeth or that Marie Antoinette said "let them eat cake," even though neither is true. Structurally, these errors are like the incorrect translations, in that they are relatively simple paired associates, with the "George Washington–wooden teeth" association needing to be updated to "George Washington–dentures made of bone and other nonwood materials." The difference is that these misconceptions are likely believed with much higher confidence than errors made when translating recently learned foreign words. In other words, we can examine either materials that elicit a range of confidence or a set designed to elicit high-confidence errors, to examine whether one's confidence in an error affects one's ability to correct it.

Intuitively, one might expect that beliefs held with high confidence would be harder to correct, as they likely reflect stronger representations in memory. However, numerous studies demonstrate that people are more likely to correct high-confidence errors than low-confidence ones (*the hypercorrection effect*; Butterfield & Metcalfe, 2001). Given feedback, most people, for example, are more likely to correct their misconception that Sydney is the capital of Australia than an error about the capital of Botswana. This pattern is observed across ages; for example, adolescents are more likely to correct high-confidence misconceptions about science, such as "The largest part of the Sahara consists of sand" and "When in a heavy thunderstorm, it is safest to lie down flat on the ground," than misconceptions that were held with lower confidence (Van Loon et al., 2015). Even young children show a higher correction rate for high-confidence errors than erroneous guesses (Marsh et al., 2012; Metcalfe & Finn, 2012).

One explanation is that people likely have greater confidence when answering questions about topics they already know something about (i.e., most Americans know more about Australia than Botswana) and that knowledge supports encoding of the feedback (for related ideas about how knowledge may support hypercorrection, see Metcalfe & Finn, 2011). A second explanation involves people's subjective feelings when faced with a large discrepancy between their confidence and accuracy – a surprising error increases attention to the feedback, with consequent benefits for memory. This claim is supported by experiments showing that people take longer to respond to a secondary tone detection task when feedback mismatches their expectations (suggesting they were distracted by the feedback; Butterfield & Metcalfe, 2006). Similarly, people are more likely to remember the color of the feedback when receiving feedback in response to correct guesses and high-confidence errors (Fazio & Marsh, 2009), supporting the hypothesis that surprising feedback directs attention toward the feedback.

The classroom implications of these results are less clear; we do not wish educators to resort to "gimmicky" feedback in order to "surprise" their students.

Rather, such results should provide reassurance that answer feedback will be sufficient even in cases where students are confident in their answers.

Misunderstandings

Misunderstanding how echolocation works is a different problem from not being able to produce the term echolocation in response to a description of how bats navigate. The first involves misunderstanding a process whereas the second is a fairly straightforward memory problem, involving forgetting of a specific term. This difference has implications for what information the feedback should convey; simply telling someone the answer works in the case of simple errors (as described in the previous two sections) but may be insufficient with more complex errors.

Many educators and researchers assume that more feedback is better. For example, educational software programs (especially Intelligent Tutoring systems) often respond to wrong answers with an explanation of *why* the answer is wrong (e.g., Graesser et al., 2005). Educators provide in-line comments on student essays, summary statements on student work, and comments on exams (e.g., Tomas, 2014). However, the experimental evidence is mixed as to whether there are added benefits from providing information beyond the correct answer (for a review, see Kulhavy & Stock, 1989; for a meta-analysis, see Bangert-Drowns et al., 1991). For example, students whose multiple-choice selections revealed misconceptions about science (such as the belief that an individual insect can become immune to pesticides) benefited as much from correct answer feedback as explanations of why their choices were wrong (i.e., that natural selection operates at the species level, not the individual level; Gilman, 1969). Similarly, it was just as effective to tell middle school children the correct answers to factual questions as to have them find the answers in their text (with line numbers to ensure they could find them; Peeck, 1979).

One issue that makes it hard to draw conclusions across studies is that the "extra information" added to the feedback takes many forms or, other times, is not specified in enough detail to evaluate. Students might receive explanations of why answers are incorrect (e.g., Kulhavy et al., 1985), reread exact texts to discover why their errors were wrong (Andre & Thieman, 1988), or be directed to look at a particular place in a text to find the correct information (Peeck, 1979), among other variations. This issue is highlighted in a study where introductory psychology students were assigned to one of four conditions, so that after each multiple-choice exam they either received no feedback, compared their answers with the correct answers written on the board, listened to the instructor discuss each question, or were directed to reread textbook passages relevant to the questions they missed (Sassenrath & Garverick, 1965). Feedback helped performance on later exam questions that tapped retention (questions that were repeated from the midterm) but it did not matter whether students received answer feedback or listened to the instructor's discussion. It is hard to draw strong conclusions about feedback content from this study, however, because it is not clear whether or not students interacted during the instructor's discussion of the feedback, nor what content was discussed. (Did the teacher focus on elaborating the correct answer by explaining why some answers were wrong? etc.) Furthermore,

even the answer feedback condition was unusual, in that it involved self-grading (comparing one's answers with those on the blackboard) and as such involved the students more than simply viewing answer feedback.

Logically, it is not clear why students would need more information than the answer to improve on a final test containing exactly the same questions as before – in such a situation, retention is required, not explanation or elaboration of knowledge. The advantage of elaborated feedback should be greatest on final tests that require going beyond retention to demonstrate a deeper understanding of the key concepts and applications of one's knowledge to novel situations. Returning to the study of introductory psychology students just discussed, the data hint at this possibility. In addition to repeating questions from the midterm, the final test included new questions that were conceptually related to some of the midterm questions, to test transfer of learning. On these transfer questions, performance was best in the discussion condition, with checking one's answers a close second. Again, it is not clear exactly what the teacher discussed but it seems reasonable to assume that a word "discussion" would involve more than just providing the answer.

We tested these ideas more directly in our own work, where students learned about complex scientific processes such as understanding how bread rises or how tornadoes form (Butler, Godbole, & Marsh, 2013). After reading the scientific texts, students took an open-ended test probing definitions of the concepts; critically, after each response, students received no feedback, were told the correct answer to the question, or received the answer in combination with an explanation that had been presented in the earlier text (Butler et al., 2013). Two days later, students took a final test that included some of the same definitional questions as on the first test, as well as novel inference questions. For example, the final test required students to name the process that facilitates gas exchange in the alveoli (definitional question) and to explain why breathing pure oxygen helps people who have trouble breathing (inference question). When faced with the same definitional questions as on the initial test, students benefited from having received feedback – but it did not matter if the feedback contained the right answer or an explanation (see Figure 17.2). In contrast, when faced with novel inference questions, students who had received explanation feedback outperformed those who had only received answer feedback. The extra information in the explanation feedback was unnecessary when the test tapped retention of answers; the explanation feedback was needed when the test required transfer of knowledge to a new context.

Conceptual Change

Even if a student understands a concept, he or she may struggle to understand how that concept interacts with other concepts or how to generalize that knowledge to a new problem. Students may sometimes have an incorrect mental model of a situation; in this case, what needs to be changed is the larger mental representation not just a specific fact or concept. We discussed earlier how feedback can be useful in correcting simple errors and here we will focus on how feedback can be used to correct a student's flawed mental representation of broad conceptual information.
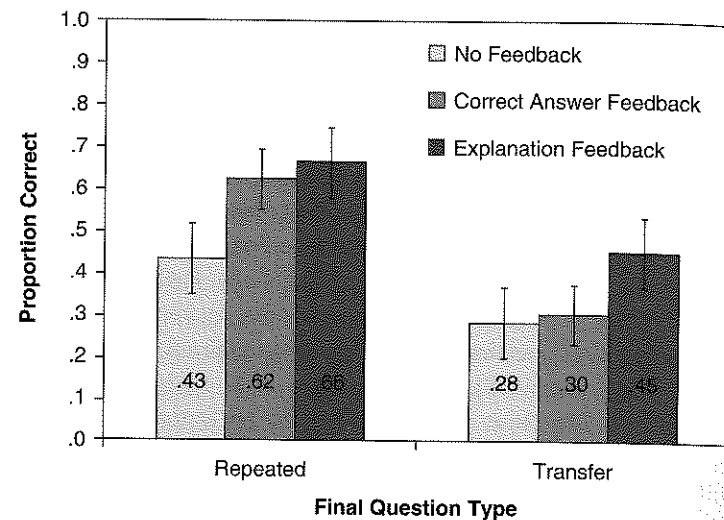
Figure 17.2 *The effect of feedback type (no feedback, correct answer feedback, explanation feedback) on repeated questions compared with transfer questions (from Butler, Godbole, & Marsh, 2013, p. 292. Adapted with permission from the American Psychological Association)*

Of course, correcting a misconception of an entire system is unlikely to occur with the kind of simple feedback described thus far – one cannot encapsulate the entirety of evolution into a single sentence, for example. Here is where the line begins to blur between learning and correction; the first time students learn about motion and friction, for example, should it be called "learning" (as it is the first school lesson on the topic) or a correction (given that the student enters the classroom with naïve incorrect beliefs about the concepts)?

For example, consider students' misunderstanding of emergent properties in science (Chi, 2005). This example requires an understanding of the differences between direct and emergent processes, so we review these first. Direct processes involve a series of sequential stages – examples include the cycles of the moon, the circulation of blood in the body, and the stages of mitosis (Chi et al., 2012). One stage is the *direct* result of an agent, prior process, or stage. In contrast, emergent processes are nonsequential and based on unconstrained, continuous action (Chi et al., 2012). The end result *emerges* from the set of actions but is not caused by any one agent or action. Examples include osmosis, heat flow, and natural selection. Consider how a student's response to a question about diffusion (an emergent process) suggests that he or she is incorrectly applying a direct process model. When describing the exchange of CO_2 and O_2 in the lungs, the student stated: "the capillaries that are in your lungs would ... let the oxygen come in through the space in its walls and then the carbon dioxide would go out ... because ... it wants to get out into a lower concentration, so all the carbon dioxide would want to go through so it would be in a lower concentration" (Chi, 2005, p. 185). The student uses sequential language, stating that oxygen *first* comes in *and then* carbon dioxide goes out. This student also

treats diffusion as an intentional process where oxygen *wants* to get in and carbon dioxide *wants* to get out. In reality, CO_2 and O_2 (and all molecules) are in constant motion (Brownian motion), moving from areas of high to low concentration – and it is these collective movements that yield diffusion, not one molecule causing another to move.

To teach emergent processing, Chi and colleagues created a module that (1) defined and differentiated the two types of processes, (2) gave everyday examples of each, and (3) prompted the student to examine how the processes played out in the everyday examples. For direct processes, students learned about wolf pack hunting and skyscraper construction; for emergent processes, students learned about schools of fish and movement in crowds. The examples were chosen to have familiar structures that students could easily understand. In the third part of the module, students identified whether the examples fit the criteria of direct versus emergent properties, as they had learned about in the beginning of the module. For example, students identified whether all agents had an equal role (indicative of an emergent process) or not (suggesting a direct process) – for example, noting that the architect's role is different from the welder (skyscraper example) but that no single fish drives the school of fish. To test the effectiveness of this module, Chi and colleagues assigned 8th and 9th grade students to complete the emergent processing module or a control science module; all students later completed a module on diffusion and took a test that tapped standard misconceptions about diffusion. Students who completed the emergent processing module (the "feedback" targeting the misconception) endorsed fewer of the misconceptions than did the students who completed the control module, although questions about the generality of this work remain.

Conceptual change is required to fix misconceptions of many complex processes, such as understanding how evolution works or why we experience seasons. Many examples can also be drawn from physics, where many students (and adults!) possess naïve beliefs about force and motion that are incorrect. For example, many people struggle to understand the concept of curvilinear motion. When people are shown the image in (the incomplete version of) Figure 17.3 of a ball on a string spun in a circle and asked to predict the path of the ball if the string were to break, they often incorrectly assume the untethered ball would continue moving in a circular pattern.

Although many people struggle with abstract physics problems, people are more likely to correctly solve such problems if they are framed within a familiar context. For example, most people have at one point in their lives engaged in a water fight, spraying others with a garden hose. If Figure 17.3 is interpreted as a person holding a garden hose (black line), people do not predict that the water will follow the curve of the hose but, instead, correctly predict that the water will shoot directly out of the hose (in a straight line, independent of the curvature of the hose). In this and other examples, the feedback involves drawing students' attention to familiar past experiences, to help them understand their mistakes in more abstract situations.

This section emphasizes the many different forms feedback can take, from simply indicating whether an answer is correct or incorrect to an entire learning module. The challenge is for educators to know what is needed, and when – points we cover in the next section.
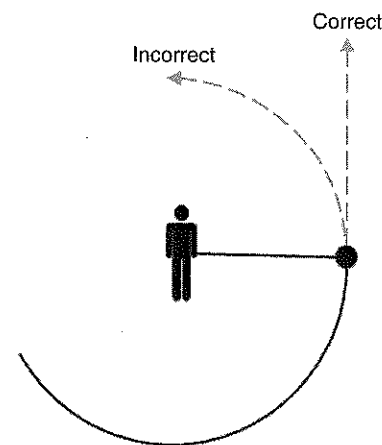
Figure 17.3 *The ball and string problem, which tests a person's understanding of curvilinear motion (after Kaiser, Jonides, & Alexander, 1986)*

This schematic shows correct and incorrect responses (dashed lines) to the ball and string problem, which tests a person's understanding of curvilinear motion.

Advice for Educators

Overview

The laboratory is obviously very different from the classroom – and even with that expectation, we were shocked at the differences we experienced when we started working in undergraduate engineering classrooms. Our experiments did not translate in the ways we expected them to – students worked together on assignments, teachers unknowingly ruined manipulations, and people cheated (Butler et al., 2014). We can only imagine the challenges involved in working in classrooms with younger students. We are not surprised that some laboratory findings do not translate to the classroom; a once-significant effect may be swamped in the classroom when the experimenter can no longer control for other factors that carry more weight.

Fortunately, it is clear that receiving feedback benefits students learning authentic educational materials, although sometimes learning is measured outside of assignments contributing to course grades. The benefits occur with such varied educational topics as soil ecology (Farragher & Szabo, 1986), army ants (Surber & Anderson, 1975), glaciers (More, 1969), the human eye (Kulhavy, Yekovich, & Dyer, 1976), and introductory psychology (e.g., Kulhavy & Anderson, 1972). Increased exposure to feedback matters; for example, the number of optional homework feedback sessions attended is correlated with final course grades (Núñez-Peña, Bono, & Suárez-Pellicioni, 2015). Quality of feedback also matters; 6th grade students' math achievement improved after teachers received training on what to include in their written feedback to homework (Elawar & Corno, 1985).

Given that feedback helps, what do educators need to know? What might surprise them? In the following section, we try to be more specific than simply advising teachers to "give feedback."

Know the Few Situations Where Feedback May Be Unnecessary

The title of this section may appear to directly contradict everything we have written thus far – so it is worth reiterating that most of the time feedback is important to provide. However, in schools, time is a precious and limited resource – critically, time spent doing one activity is at the cost of another. It remains an open question how educators should decide to spend valuable class time, given that the time used to distribute feedback could be used for other learning activities known to benefit learning, such as retrieving information from memory (i.e., flashcards, quizzes; McDaniel et al., 2011) or writing to learn activities (for a meta-analysis, see Bangert-Drowns, Hurley, & Wilkinson, 2004; see also Klein and Van Dijk, Chapter 8, this volume).

One situation was alluded to earlier in this chapter: Feedback has relatively little effect on the maintenance of correct answers, *unless* they were guesses (Butler & Roediger, 2008). In contrast, feedback is absolutely crucial for error correction and should not be skipped. The logical extension of these findings is that feedback may not be necessary if student performance is uniformly high. This point was captured in a laboratory study where undergraduates received a fixed amount of time to learn two lists of Swahili–English translations (Hays, Kornell, & Bjork, 2010). After an initial study phase, students completed a series of test-feedback trials (akin to going through a pack of flashcards repeatedly). Feedback viewing was required for one list (every flashcard had to be turned over), whereas it could be skipped for the other list, as determined by either the computer or the learner. In this case, skipping feedback (whether determined by the computer or the student) meant that students spent more time practicing retrieval (as opposed to reading feedback), which led to more correct translations a day later (reflecting the known memorial benefits of retrieving information from memory; e.g., Roediger & Butler, 2011). It should be noted, however, that participants' judgments about whether or not to skip feedback viewing were excellent, with 85 percent of their feedback choices lining up with their actual performance. As covered in the next section, students are not always so calibrated at judging their need for feedback, meaning it is likely safer for the teacher to make the decision that feedback is unnecessary.

Ensure That Students Look at the Feedback

In the section "Errors Made With Confidence", we argued that surprising feedback was better attended to, with consequent benefits for later correction (Fazio & Marsh, 2009). Yet it is difficult to directly generalize these results to educational practice – not all feedback surprises the learner and, in many cases, the teacher has no idea if students even looked at the feedback, let alone understood it. In the laboratory, we can address these problems, requiring feedback to stay on the computer screen for

a set period of time, instructing students to respond after reading it, or asking students to make a judgment that shows they processed it for meaning (e.g., Lhyle & Kulhavy, 1987). However, educators normally do not have the option of forcing students to spend a set amount of time reading feedback, given that time spent on review means less time for new learning activities.

Furthermore, attention is not guaranteed, even in a relatively captive classroom given that laptops distract (Fried, 2008) and minds wander (see review by Szpunar, Moulton, & Schacter, 2013). The laboratory likely underestimates the challenges of directing a learner's attention to processing feedback in authentic educational settings, where feedback processing is often left to the discretion of students. We have all seen students recycle their commented papers as they walk out the classroom door, or experienced empty office hours when no students stop by to view their exams (which were not returned in order to protect a bank of test questions). Existing data confirm these anecdotal impressions. For example, one set of researchers took advantage of the practice of allowing third-year medical students to submit a self-addressed stamped envelope (SASE) so feedback on their essay could be mailed to them. This essay was a five-page review piece and a passing grade was required to advance to the next grade. Unfortunately, less than half of the students provided the SASE, meaning most of the students did not receive any feedback on their written work (Sinclair & Cleland, 2007). Perhaps most discouraging, students who likely needed the feedback the most (the ones with lower course grades) were the least likely to provide an envelope so that the feedback could be mailed.

In our study in an engineering classroom, we used an online homework system that allowed us to automatically collect records of whether or not individual students clicked on problem-by-problem feedback (Mullet et al., 2014). This system also made it possible to require feedback viewing, if desired. In one section of the course, feedback viewing was required and counted toward the class grade. In the other section, feedback viewing was optional, as is the norm in most college courses. The results were striking: When feedback viewing was required, 94 percent of students clicked on the links. In contrast, students in the feedback-optional section only clicked on the feedback links for 47 percent of the problems. For a given problem, students viewed the feedback sooner and more frequently in the feedback-required section than in the optional section. These different behaviors were associated with differences in performance on the course exams – students who had been required to view the feedback answered 10 percent more exam questions correctly than did the students for whom feedback viewing was optional.

Why did students sometimes fail to access the feedback provided to them? One possibility is that students simply punt on tasks that do not contribute directly to their grades. A second possibility is that students may not always be calibrated enough to know when to look at the feedback. It is entirely possible that the students in the engineering classroom study believed they looked at the feedback when they really needed it – which would suggest a metacognitive problem, not a laziness problem. Both are likely involved, but here we focus on the metacognitive issue.

Correctly skipping feedback depends critically on people's awareness of when they made mistakes. In the Hays and colleagues (2010) study discussed in the

previous section, students benefited from the ability to skip feedback – but those students were very good at knowing which Swahili words they could not translate, with 85 percent of their feedback choices lining up with their actual performance. We cannot assume similar calibration of learning with more complex materials. It may be obvious when one does not know the translation of a foreign word but much harder to judge the quality of one's essay or whether a math problem was solved correctly. Second, students in that study likely benefited from skipping feedback because they replaced that time with another learning activity – skipping feedback on some trials allowed them to spend more time retrieving other information. We do not have any data on this point but we doubt that the students in our feedback-optional section used the time they saved on some other additional activity for the class.

Ensure That Students Process the Feedback Correctly

Our engineering students benefited from clicking on the feedback links – but it should be noted that we do not know if they actually read the content, whether they thought about it, and so on. In some instances, additional steps may be necessary to ensure that students actually process the feedback. For example, in one study, experimenters attempted to leverage the known benefits of active processing (e.g., Slamecka & Katsaiti, 1987) by requiring students to unscramble the words in the feedback message (that is, the correct multiple-choice alternative was presented in scrambled format). This strategy only helped when the experimenters added a task to make sure that students actually processed the feedback for meaning, in this case, by requiring them to write out the correct unscrambled version of the feedback (Lhyle & Kulhavy, 1987).

More generally, feedback will fail if students do not understand it or fail to understand how it contradicts their own answer. This may be particularly important in situations where people are self-grading or peer-grading – if they do not apply a rubric correctly, they will not realize an error has been made. In one study, students were asked to define terms such as "the fundamental attribution error" and then compare their answers to correct definitions (Rawson & Dunlosky, 2007). The critical finding is that sometimes students could not understand how their incorrect answer differed from the correct one. A similar finding has been reported with example generation, with students asked to generate examples of concepts failing to understand how their answers fell short when provided with feedback (Zamary, Rawson, & Dunlosky, 2016).

Don't Assume That Corrections Are Permanent

There is no "mastery" level that guarantees information will always be known – for example, we have all experienced tip-of-the-tongue states whereby something we know is currently "out of reach." More generally, forgetting increases as the cues in the environment change and as time passes since learning and using a piece of information. The most well-known educational example of this problem is the so-

called summer learning loss whereby students appear to lose large amounts of information over summer vacation (Cooper et al., 1996). One issue involves how to reactivate that previously learned information; another problem involves the potential resurgence of an earlier error.

First, consider the problem whereby a student has forgotten the meanings of core concepts from a class. In at least some cases, the simple solution involves simple reexposure to the material – the same way a tip-of-the-tongue state is often resolved when someone takes pity on the speaker and fills in the missing word. In our own work, we have shown that a multiple-choice test can serve the same function; after knowledge about the Treaty of Versailles is forgotten, solving the multiple-choice question "What peace treaty ended World War I?" is sufficient to reactivate it (while also providing formative assessment to the teacher). One problem involves identifying which information should be revisited; when we investigated these issues in a pharmacology classroom, we relied on the instructor to identify foundational material that students were supposed to know from prerequisite coursework (Butler et al., under review). For three of the course's six units, students answered multiple-choice questions on related foundational material (which units were assigned to the intervention versus control were counterbalanced across subjects). Pretesting indicated that students were unable to produce 75 percent of the foundational material at the start of the course (even though it all had been covered in prerequisites for the course); multiple-choice testing (without feedback) led to significant improvement on a later test of that foundational material. To our knowledge, this is the first demonstration of stabilizing access to foundational knowledge in a classroom setting but there is still much to be learned about how to identify which knowledge needs to be reactivated and how often – a point we return to at the end of this chapter.

The second problem involves the resurgence of errors. Spontaneous recovery of undesirable behaviors is common – the drinker who abstained from alcohol can "fall off the wagon," a frequent flyer may become anxious after numerous smooth flights, and long-debunked beliefs (i.e., that the world is flat) become popular again. In education, this problem is particularly problematic when dealing with confidently held misconceptions. Earlier in this chapter, we discussed how people are more likely to correct high-confidence errors such as "Sydney is the capital of Australia" than erroneous guesses, given feedback. This result is surprising because confidently held errors would be expected to be harder to correct – and yet increased attention to the feedback leads to a higher correction rate. But what happens after time has passed? Forgetting of the recent event (the feedback) will occur at a faster rate than forgetting of the misconception (this is Jost's Law; Jost, 1897) – meaning that, at some point in time, the misconception will be stronger in memory than the feedback. A week later, it was errors such as "Sydney is the capital of Australia" that were most likely to reappear (Butler, Fazio, & Marsh, 2011).

Finally, vestiges of misconceptions may remain even though students show knowledge of the correct information. For example, young children often believe that the ability to move is a prerequisite for life, a belief that is at odds with plants being alive. The remnant of this belief shows up when undergraduates are asked to

quickly classify a series of items as living versus nonliving; they are slower and less accurate to classify plants than animals, even after controlling for word frequency (Goldberg & Thompson-Schill, 2009). Even PhDs in biology show a similar effect when told to make their decisions quickly, despite having spent an average of twenty years as biology faculty. Under the right circumstances, an ingrained false belief may coexist with its correction. Similarly, even though most people know that density determines whether an object floats or sinks in water, they are slower to make decisions when mass and density are not positively correlated, reflecting the childhood tendency to relate buoyancy and size (Potvin et al., 2015).

Student Preferences Do Not Always Align with the Best Learning Strategies

Student beliefs about learning are notoriously incorrect. For example, many students prefer to use inefficient study strategies such as rereading their textbooks and rewriting their notes instead of engaging in self-testing (Karpicke, Butler, & Roediger, 2009). Cramming is another popular, yet ineffective, study technique often used by students. In fact, students believe they learn more when their study sessions are massed together even though this is not true – they actually learn more when they spread out their study sessions (Kornell, 2009). These metacognitive illusions likely occur because easy practice feels good, even if it does not promote long-term learning. That is, rereading a text is easier than reading it for the first time, and a translation is easily retrieved if I just tested myself with the same flashcard two trials ago – tricking the learner into believing they are good study strategies (see Bjork & Schmidt, 1992). It matters what students like, as their preferences are correlated with their teacher ratings (Beleche, Fairris, & Marks, 2012) which in turn often play a role in how teachers are evaluated in their jobs.

Students tend to like feedback that includes written comments, although the need for them likely varies depending on the nature of the comment and the type of to-be-learned information. While students value written feedback, not all written feedback is viewed as helpful (Weaver, 2006). A common complaint of students is that the written feedback they received on their assignments was unclear, confusing, and needed to include more details in order for students to understand how to improve (Ferguson, 2011; Walker, 2009). Earlier we discussed how feedback need not always contain an explanation of why a particular answer is right or wrong – but students like to receive feedback that explains their grade. For instance, one student complained that there were "not enough comments to justify the grade given" (Ferguson, 2011, p. 57). Specifically, students prefer feedback that contains comments on the quality of their main ideas rather than specific details. Unsurprisingly, students like to see positive comments to build their confidence included with more critical feedback indicating how to improve their work (Ferguson, 2011).

A second misconception of students is that feedback should be provided as soon as possible (of course, this preference likely also reflects their desire to know their grades). This belief is not limited to students; it occurs in contemporary educational programs and is often considered a positive feature of an assessment if feedback can

be immediate – Coursera touts immediate feedback to student responses in massive open online courses (MOOCs) and teachers can purchase the Immediate Feedback Assessment Technique (IF-AT) testing system (where students uncover a star if they scratch off the correct answer to a multiple-choice question; Epstein, Epstein, & Brosvic, 2001; Epstein & Brosvic, 2002). This belief traces back to B. F. Skinner, who showed that animals required immediate reinforcement to learn an association between a lever press and a food reward (see review by Renner, 1964). However, the results with humans are actually quite mixed, with some studies (mostly classroom studies) showing a benefit of immediate feedback whereas others (mostly laboratory studies) show a benefit of delaying feedback (for meta-analysis, see Kulik & Kulik, 1988). Despite the ambiguous success of immediate feedback, students show a strong preference of using the IF-AT where they receive immediate corrective feedback over traditional multiple-choice testing (Dibattista, Mitterer, & Gosse, 2004). On written assignments, university students preferred feedback to be returned in two to three weeks, as long as it was before the start of the next assignment (Ferguson, 2011).

In our own work on this issue, we used an online homework system that allowed us to carefully control the timing of feedback (Mullet et al., 2014). Students in upper-level undergraduate engineering classes received identical feedback on their homework assignments; the only difference was whether the feedback was delivered immediately after the homework deadline or delayed by one week. Grades on course exams were higher following delayed feedback, even after controlling statistically for the shorter retention interval. However, students failed to recognize the benefits of delayed feedback. When asked which feedback schedule they preferred and which one was more effective, the overwhelming majority of students reported a strong preference for immediate feedback. This was true regardless of whether they experienced both schedules of feedback within their course (a within-subjects design) or only experienced one of the two schedules (a between-subjects design). Thus, there was a metacognitive disconnect between the feedback timing schedule that students preferred and what actually helped them to learn.

Given the general importance of student ratings in teacher evaluations, we understand why teachers might be loath to implement a strategy that promotes learning but is almost universally disliked. One possibility is to explain upfront to students the reasons for one's choices (which is possible in practice, albeit not in a controlled experiment) – in teaching, we have found that students interpret delayed feedback as evidence that the teacher does not care or is procrastinating; it may not be possible to completely manage this impression but explaining the reasoning will not hurt and has the potential to help.

Consider Students' Preexisting Beliefs and Motivations to Change

From the instructor's perspective, feedback consists of relatively neutral information (other than the negative affect associated with making a mistake). However, learners enter the classroom with preexisting beliefs and differ in their motivation to change them. No one is surprised by a lack of change when one simply tells a climate change

doubter to believe science on global warming – a different approach is needed. Even in a study with relatively neutral statements (e.g., telling students that they were wrong for believing that bulls become enraged by the color red), students indicated that they did not believe all of the feedback. That is, the more confident students were in their initial responses, the less likely they were to believe the feedback, as rated on a 0 (do not believe the feedback at all) to 100 (absolutely believe the feedback) scale (Rich et al., 2017). Belief in the feedback, in turn, was related to later corrections of the errors.

This type of situation is also one where explanation feedback can help. It makes sense that when someone is motivated to believe something, he or she will need more evidence to reject it. Accordingly, participants who received both the right answer and the explanation behind it were more successful at correcting their misconceptions than participants who only received feedback with the correct answer (Rich et al., 2017). That is, receiving the following feedback, "The color red does not enrage bulls (the correct answer) because bulls do not see the color red, and, instead, attack because they perceive the matador as a threat (the explanation)" (Rich et al., 2017, p. 492), is more likely to increase your belief in the feedback and help you correct your misconception than only receiving "The color red does not enrage bulls (the correct answer)" (Rich et al., 2017, p. 492). When tackling students' mistaken beliefs, feedback is only helpful in correcting misconceptions if it is believed. Providing explanations in addition to the correct answer is a good way to refute the misconception and help students accept the feedback as true.

Conclusions

Open Questions

One open question involves the value of personalized feedback. Personalized learning is, in general, a hot topic in education – it is very appealing to think about adjusting the curriculum and feedback to a particular learner. While there are many commercial products touting personalized learning, in most cases we know little about the nature of the underlying algorithms. From the academic perspective, the best evidence for personalized learning comes from work looking at the effects of different practice schedules on the retention of Spanish vocabulary words (Lindsey et al., 2014). Grade 8 students practiced Spanish vocabulary via an online flashcard (retrieval followed by feedback) tutoring system called the Colorado Optimized Language Tutor (COLT). In a standard practice condition, practice was massed – one-third of the words were assigned to be translated chapter-by-chapter, so that older chapters were never revisited as the class progressed through the book. In a spaced condition, the scheduler had one-third of the words revisited later in the course – practice was spread out over time. In a third condition, analytics were used to individually determine the practice schedule for one-third of the words; the analytics drew on both the learner's data and a large amount of data about past learners to make predictions about which items needed the most practice. Briefly, the

results showed that children learned the words in the personalized spacing condition best, followed by those in the spaced condition, with the worst retention of words in the standard practice condition (especially those that had occurred early in the course).

Feedback was provided in all three COLT learning conditions, so the current data do not tell us anything about how feedback should be scheduled. However, some evidence suggests that the temporal spacing of feedback is similar to that of retrieval (Smith & Kimball, 2010). That is, a fair amount of evidence suggests that the interval between learning opportunities should be between 5 percent and 20 percent of the desired retention interval (Cepeda et al., 2008) – and this same formula works for the timing of feedback.

Final Thoughts

Vague advice to educators can be harmful rather than helpful if principles are implemented in a way that changes the processing involved. One concern is that we are simply asking too much of teachers, who are already often following mandated curricula – especially since we suspect many of the chapters in this volume are providing other pieces of advice. However, we encourage teachers to keep in mind (1) the benefits of delaying feedback and (2) the resurgence of errors over time. To the extent that new topics build on old ones, there may be a natural spacing of feedback over time. The teacher can also watch for any indicators that students are regressing, which could trigger a need for review. More generally, we encourage teachers to think more flexibly about the definition of feedback. Feedback could take the form of a multiple-choice quiz (Cantor et al., 2015), a student presentation, or an in-class review game – it does not have to take the form of responses to a test or other graded assignment.

References

Andre, T. & Thieman, A. (1988). Level of adjunct question, type of feedback, and learning concepts by reading. *Contemporary Educational Psychology, 13*(3), 296–307.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29–58.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review, 31*(5), 709–719.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*(1), 417–444.

Butler, A. C., Black-Maier, A. C., Campbell, K., Marsh, E. J., & Persky, A. M. (under review). Stabilizing access to marginal knowledge in a classroom setting.

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin and Review, 18*(6), 1238–1244.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*(2), 290–298.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. III (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273–281.

Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. *Educational Psychology Review, 26*(2), 331–340.

Butler, A. C. & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory and Cognition, 36*(3), 604–616.

Butterfield, B. & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1491–1494.

(2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*(1), 69–84.

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory and Cognition, 43*(2), 193–205. http://dx.doi.org/10.3758/s13421-014-0462-6

Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41*(10), 1123–1130.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095–1102.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14*(2), 161–199.

Chi, M. T., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive science, 36*(1), 1–61.

Clancy, M. (2004). Misconceptions and attitudes that interfere with learning to program. In S. Fincher & M. Petre (eds.), *Computer science education research* (pp. 85–100). CRC Press.

Clarke, L. K. (1988). Invented versus traditional spelling in first graders' writings: Effects on learning to spell and read. *Research in the Teaching of English, 22*(3), 281–309.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227–268.

Dihatista, D., Mitterer, J. O., & Gosse, L. (2004). Acceptance by undergraduates of the immediate feedback assessment technique for multiple-choice testing. *Teaching in Higher Education, 9*(1), 17–28.

Elawar, M. C. & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162–173. https://doi.org/10.1037/0022-0663.77.2.162

Epstein, M. L. & Brosvic, G. M. (2002). Students prefer the immediate feedback assessment technique. *Psychological Reports*, *90*(3, suppl.), 1136–1138.

Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports*, *88*(3), 889–894.

Farragher, P. & Szabo, M. (1986). Learning environmental science from text aided by a diagnostic and prescriptive instructional strategy. *Journal of Research in Science Teaching*, *23*(6), 557–569.

Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L., III (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory and Cognition*, *38*(4), 407–418. http://dx.doi.org/10.3758/MC.38.4.407

Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory*, *18*(3), 335–350.

Fazio, L. K. & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*(1), 88–92.

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment and Evaluation in Higher Education*, *36*(1), 51–62. https://doi.org/10.1080/02602930903197883

Fried, C. B. (2008). In-class laptop use and its effects on student learning. *Computers and Education*, *50*(3), 906–914.

Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology*, *60*(6, Pt. 1), 503–508.

Goldberg, R. F. & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*(4), 480–487.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612–618.

Hart, J. T. (1967). Second-try recall, recognition, and the memory-monitoring process. *Journal of Educational Psychology*, *58*(4), 193–197.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York: Routledge.

  (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. https://doi.org/10.1037/stl0000021

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin and Review*, *17*(6), 797–801.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abha"ngigkeit von der Verteilung der Wiederholungen [The strength of associations in their dependence on the distribution of repetitions]. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane*, *16*, 436–472.

Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory and Cognition*, *14*(4), 308–312.

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479.

Klahr, D. & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology*, *20*(3), 362–404.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Kulhavy, R. W. & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, *63*(5), 505–512. https://doi.org/10.1037/h0033243

Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.

Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*, *10*(3), 285–291.

Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology*, *68*(5), 522–528.

Kulik, J. & Kulik, C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97. www.jstor.org/stable/1170349

Lhyle, K. G. & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology*, *79*(3), 320.

Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students' long-term knowledge retention through personalized review. *Psychological Science*, *25*(3), 639–647.

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers and Education*, *95*, 85–98.

Marsh, E. J., Fazio, L. K., & Goswick, A. E. (2012). Memorial consequences of testing school-aged children. *Memory*, *20*(8), 899–906.

Marsh, E. J., Lozito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory*, *20*(6), 645–653.

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin and Review*, *14*(2), 194–199.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III, (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414.

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*, *37*(2), 437–448.

  (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction*, *22*(4), 253–261.

More, A. J. (1969). Delay of feedback and the acquisition and retention of verbal materials in the classroom. *Journal of Educational Psychology*, *60*(5), 339–342. https://doi.org/10.1037/h0028318

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, *3*(3), 222–229.

Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education, 69*(3), 191–196.

Núñez-Peña, M. I., Bono, R., & Suárez-Pellicioni, M. (2015). Feedback on students' performance: A possible way of reducing the negative effect of math anxiety in higher education. *International Journal of Educational Research, 70*, 80–87.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3–8.

Peeck, J. (1979). Effects of differential feedback on the answering of two types of questions by fifth-and sixth-graders. *British Journal of Educational Psychology, 49*(1), 87–92.

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education, 13*(1), 21–43.

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15*(1), 27–35.

Rawson, K. A. & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology, 19*(4–5), 559–579.

Renner, K. E. (1964). Delay of reinforcement: A historical review. *Psychological Bulletin, 61*(5), 341–361. https://doi.org/10.1037/h0048335

Rich, P. R., Van Loon, M. H., Dunlosky, J., & Zaragoza, M. S. (2017). Belief in corrective feedback for common misconceptions: Implications for knowledge revision. *Journal of Experimental Psychology: Learning, Memory, And Cognition, 43*(3), 492–501. https://doi.org/10.1037/xlm0000322

Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Sassenrath, J. M. & Garverick, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology, 56*(5), 259–263.

Schmidt, R. & Bjork, R. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217.

Sinclair, H. K. & Cleland, J. A. (2007). Undergraduate medical students: Who seeks formative feedback? *Medical Education, 41*(6), 580–582.

Skinner, B. F. (1953). *Science and human behavior.* New York: Simon and Schuster.

Slamecka, N. J. & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26*(6), 589–607.

Smith, T. A. & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 80–95. https://doi.org/10.1037/a0017407

Surber, J. R. & Anderson, R. C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology, 67*(2), 170–173. https://doi.org/10.1037/h0077003

Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind wandering and education: From the classroom to online learning. *Frontiers in Psychology, 4*, 495.

Terrace, H. S. (1963). Errorless transfer of a discrimination across two continua. *Journal of the Experimental Analysis of Behavior, 6*(2), 223–232.

Tomas, C. (2014). Marking and feedback provision on essay-based coursework: A process perspective. *Assessment and Evaluation in Higher Education, 39*(5), 611–624. https://doi.org/10.1080/02602938.2013.860078

Van Loon, M. H., Dunlosky, J., Van Gog, T., Van Merriënboer, J. J., & De Bruin, A. B. (2015). Refutations in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology, 42*, 39–48.

Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment and Evaluation in Higher Education, 34*(1), 67–78. https://doi.org/10.1080/02602930801895752

Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education, 31*(3), 379–394. https://doi.org/10.1080/02602930500353061

Yeager, D. S. & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302–314.

Zamary, A., Rawson, K. A., & Dunlosky, J. (2016). How accurately can students evaluate the quality of self-generated examples of declarative concepts? Not well, and feedback does not help. *Learning and Instruction, 46*, 12–20.