



Providing Learners with Feedback—Part 2:

Peer-reviewed research compiled for
training, education, and e-learning.



WORK-LEARNING RESEARCH, Inc.

A Work-Learning Research Publication. May 2008.

© Copyright 2008 by Will Thalheimer. All rights are reserved.

You may share this report as long as you share it intact and whole and don't charge a price for it or include it in any priced event or product.

This is a two-part report. Part 1 provides background and recommendations. Part 2 focuses on research support.

How to cite this report using APA style:

Thalheimer, W. (2008, May). *Providing Learners with Feedback—Part 2: Peer-reviewed research compiled for training, education, and e-learning*. Retrieved November 31, 2008, from <http://www.work-learning.com/catalog/>

Obviously, you should substitute the date on which the document was downloaded for the fictitious November 31 date.

Published May 2008

The Author: Will Thalheimer is a learning-and-performance consultant and research psychologist specializing in learning, cognition, memory, and performance. Dr. Thalheimer has worked in the workplace learning field, beginning in 1985, as an instructional designer, simulation architect, project manager, product leader, trainer, consultant, and researcher. He has a PhD from Columbia University and an MBA from Drexel University. He founded Work-Learning Research in 1998 to help client organizations create learning environments to maximize performance and help learning professionals—instructional designers, e-learning developers, trainers, performance consultants, talent managers, and chief learning officers—utilize research-based knowledge to build effective learning-and-performance solutions.

Dr. Thalheimer can be contacted at info@work-learning.com or at the Work-Learning Research phone number below for inquiries about learning audits, workshops, speaking engagements, high-level instructional design, and consulting on e-learning, classroom training, learning measurement, learning strategy, and business strategy for learning.

Work-Learning Research, Inc.
Somerville, Massachusetts USA
(617) 666-9637

www.work-learning.com

All information contained in this document is copyrighted.
© Copyright 2008 by Will Thalheimer. All rights are reserved.

Sharing Rules: You may copy and distribute this document as long as you share it intact and whole and don't charge a price for it or include it in any priced event or product. (this includes generating good will that might be translated in the future into tangible financial benefits).

We do not receive a salary for translating research into practical recommendations. If you abuse the spirit of the “Sharing Rules” describe above, you are not only stealing, you are limiting our ability to continue to support you and others in the learning-and-performance field!

Introduction

Hi. I'm Dr. Will Thalheimer, a consultant and researcher specializing in learning fundamentals, instructional design, performance improvement, learning measurement, and workplace learning. I help people create more effective learning interventions by building bridges between learning research and learning practice. There is wisdom in both camps, but only by integrating the two can we maximize our learning outcomes.

I've been researching the topic of feedback for almost 10 years (off and on), studying research articles from the world's preeminent refereed journals, translating that research to provide practical wisdom for learning professionals, and compiling what I've learned to make it available to others. Although I've shared my tentative findings with a few others in the past, this document represents my first published research-to-practice report dealing directly with the topic of feedback.

The feedback literature is probably the most difficult learning-research area I have had the pleasure to research and translate. In short, the research landscape on feedback is a war zone, complete with countless shattered structures, broken windows, and unexploded ordnance. Those poetically inclined can parse these symbols for their full meaning, but let me just say this: as Hemingway and others found some wisdom and solace in war, the feedback literature has its own bounty of truth to tell, even in the midst of its unfolding story (as new research continues to become available).

This report is designed to help you improve your feedback methods. I certainly won't claim to have all the answers, nor do I think simple recipes are available in dealing with feedback. I do believe strongly, however, that all of us can improve our feedback methods substantially, and by so doing improve the practice of education and training.

I would like to thank Questionmark for agreeing in advance of my final writing efforts to license this report for the benefit of their clients. Questionmark is available on the Web at www.questionmark.com and by phone at 800-863-3950 (North America), +44 (0)20 7263 7575 (United Kingdom) or +32 2 298 02 01 (Europe).

This Report's Design

This is a research-to-practice report, written in two parts. Part 1 is written for a general audience. It provides perspective on feedback's place in learning and it makes specific recommendations for practice. It also includes a two-page summary of the practical recommendations. Part 2 contains a research review that supports the recommendations made in Part 1. Research references are included in Part 2.

Because some of the concepts and terminology in this report may introduce you to new paradigms, you can expect that they will seem foreign at first. As their power becomes evident—as you build new mental models of these concepts—the initial fog will clear.

Contents

Introduction.....	3
This Report’s Design	3
Research Review.....	5
The Power of Retrieval Practice	5
Is Feedback Effective?.....	6
Feedback on Correct versus Incorrect Answers.....	8
Feedback Enables Correct Retrieval Practice and Follow-up Retrieval Practice Produces Benefits.....	12
Should Learners Get Feedback Only After Making a Response?	13
Timing of Feedback—Which is Better, Immediate or Delayed Feedback?	14
Research Reviews on Feedback Timing are Not Definitive.....	16
Two New Research Threads on the Timing of Feedback.....	17
The Distinction Between Knowledge, Marginal Knowledge, and Submarginal Knowledge.....	20
Feedback’s Effects Can Be Minimized.....	21
Feedback for Physical Movement Skills.....	21
Other Issues to Consider for Feedback Design.....	24
The Specificity of Feedback	24
Getting Learners to Pay Attention to the Feedback	24
Using Confidence Ratings to Augment Feedback	25
Using the Answer-Until-Correct Methodology	26
The Distinction between Building Understanding and Supporting Retrieval/Fluency	28
Complexity of Feedback Context Makes Research Difficult:	29
A Message for Researchers.....	29
Considering All Phases of the Feedback Context.....	29
Critical Variables to be Considered	30
Priorities for Further Research.....	32
Summarizing Principles and Recommendations (copied from Part 1).....	34
Reference List	36

Research Review

Over the years as I've attempted to convey research-based recommendations to learning professionals like you, I've found that most people hate reading dense research-like prose. I don't blame you. It's hard work. Also, it's nearly impossible to read an actual research article and know what to do. First, you have to understand the article. Second, you have to know how that particular article fits into the research thread that has been created over the years. Third, you have to know how and how much to discount the findings because of methodological limitations. Fourth, you have to consider how the research article generalizes to your particular situation. Fifth, you have to be skeptical of the researcher's conclusions because they don't always follow logically from the data. Finally, you have to freakin' take a vacation to recover.

If you've read Part 1 research-to-practice report, thank you and congratulations. For most people, if you put what you've learned into practice, you'll significantly improve your feedback results. The rest of this document provides a research review. It's easier to read than an actual research article, but not much easier. It's better than an actual research article in that it reviews a breadth of research, but it's limited by the distillation of detail that reviews require. It will give you a flavor for the research, provide perspective on the strength of research-based recommendations, and provide repetition and reinforcement for what you've already garnered. Go at it with gusto!

The Power of Retrieval Practice

Many researchers have found that testing people after they have learned something—in other words providing them with retrieval practice—is more effective in promoting their later performance than providing additional time to study the material (Roediger & Karpicke, 2006b; Butler & Roediger, 2007; Nungester & Duchastel, 1982; Hogan & Kintsch, 1971; Cuddy & Jacoby, 1982; Kuo & Hirshman, 1996; Izawa, 1992; Allen, Mahler, & Estes, 1969; Jones, 1923-1924).

Providing more testing is better than providing less testing up to a certain point of diminishing returns (Izawa, 1992; Allen, Mahler, & Estes, 1969; Modigliani & Hedges, 1987; Glover, 1989; Bangert-Drowns, Kulik, & Kulik, 1988; Crooks, 1988). To put an exclamation point on this, Bahrck, Bahrck, Bahrck, and Bahrck (1993) found that 26 tests with feedback outperformed 13 tests with feedback, producing 33% more retention on subsequent tests given several years later. Bahrck (1979) found that six tests with feedback outperformed three tests with feedback whether these tests were repeated immediately, after one day, or after 30 days by 106%, 34%, and 32% respectively. Similarly, repeating questions within a single test can improve performance somewhat on later tests (Toppino & Brochin, 1989; Cuddy & Jacoby, 1982).

The power of retrieval is also evident in the way testing can affect future learning. Learners may change their attention or cognitive processing as they work on new

material, as has been found in the adjunct question research (e.g., Rickards, 1979; Rothkopf, 1966; Frase, 1967). For example, Rothkopf and Bisbicos (1967) found that certain categories of questions given to learners on previous unrelated material affected what they learned from subsequent learning opportunities. Learners did better on questions about proper names and numerical data when they had previously gotten similar questions on previously read text passages. Similarly, Sagerman and Mayer (1987) found that learners did better on verbatim questions when they had previously gotten verbatim questions and conceptual questions when they had previously gotten conceptual questions¹. It should be noted that these “forward” attentional effects may have a very short shelf-life. For example, in both the Rothkopf-Bisbicos and the Sagerman-Mayer studies, learners got the second set of learning materials right after the initial set of test questions.

The potency of retrieval has been noted in several recent research reviews (Roediger & Karpicke, 2006a; Pashler, Rohrer, Cepeda, & Carpenter, 2007; Bjork, 1988; Crooks, 1988) and has a long history of experimental validation, starting early in the last century (for example, Gates, 1917; Jones, 1923-1924; Spitzer, 1929).

Is Feedback Effective?

Let’s start with the basic question. Does feedback create learning that aids later retrieval? The answer is yes. In general, feedback produces beneficial effects. Many researchers who have reviewed numerous research studies are so sure of feedback’s effects that they simply assume feedback improves learning and go on to discuss other variables that affect feedback’s impact (Kulik & Kulik, 1988; Mory, 1992, 2004; Kulhavy, 1977; Kulhavy & Stock, 1989; Keenan & Langer, 1993). Bangert-Drowns, Kulik, Kulik, and Morgan (1991) did a meta-analysis of research studies and found that feedback generally aided performance, especially when the studies were methodologically sound².

As is typical in the research, not all studies have found benefits from feedback and some have found harm. Mory’s (2004) review of the feedback literature documents many inconsistencies. Kluger and Denisi (1996), reviewing both instructional and non-instructional feedback research, were very strong in arguing that feedback is not always effective. However, we must take their advice with skepticism given that their meta-

¹ This phenomenon can have detrimental effects when we continually test learners on meaningless fragments of facts, figures, and folderol. In training situations, we typically want learners to have broad understandings first, and knowledge of facts second. In educational situations, we want learners to learn how to think, observe, test their hypotheses, learn general themes, etc. The preponderance of rote testing only serves to produce people who can remember facts. It’s not the people who do well on quiz shows like Jeopardy that produce innovations, create art, and develop provocative ideas, it’s people who use their minds to understand. It’s not the memorizers who move the world forward, so why do we try to create them with our tests?

² Effect sizes reported by Bangert-Drowns, Kulik, Kulik, and Morgan (1991) were a moderate 0.46 when controlling for “presearch availability” (i.e., looking ahead to get the answer before answering the test question), a term coined by Kulhavy (1977) and a phenomenon brought to light by Anderson, Kulhavy, & Andre (1971).

analytic data included a large number of non-instructional situations that are unlikely to be relevant to formal learning situations.

Because the literature on feedback encompasses so many different research methodologies, and because many of these methodologies are inappropriate or irrelevant to the topic of feedback³, it is most useful to explore original research studies that are appropriately designed instead of relying solely on research reviews.

Studies that compare feedback to not giving feedback generally find improvements with feedback—sometimes small but oftentimes quite substantial improvements (Butler, Karpicke, & Roediger (in press for 2008a, submitted for 2008b); Pashler, Cepeda, Wixted, & Rohrer, 2005; Brosvic, Epstein, Dihoff, & Cook, 2005; Kang, McDermott, & Roediger, 2007; Karraker, 1967; Kulhavy & Anderson, 1972; Kulhavy, Yekovich, & Dyer, 1976; Surber & Anderson, 1975; Sturges, 1978; Clariana, Ross, & Morrison, 1991; Webb, Stock, & McCarthy, 1994).

For example, Pashler, Cepeda, Wixted, and Rohrer (2005, p. 6) found that “supplying the correct answer after an incorrect response increased 1-week retention by 494% as compared with the no-feedback condition.” Kang, McDermott, and Roediger (2007) found that feedback on incorrect answers was better than no feedback by about 77%. Surber & Anderson (1975) found improvements due to feedback by an average of 47% for questions that were wrong. Note how all three of these results were found when feedback was given after incorrect answers. As we’ll discuss in the next section, feedback has its most profound effect for incorrect answers and much less of an effect when provided for correct answers. For example, Sturges (1978) found average improvements due to feedback overall (for both correct and incorrect answers) at 11%, but improvements when feedback was given for incorrect answers at 64%.

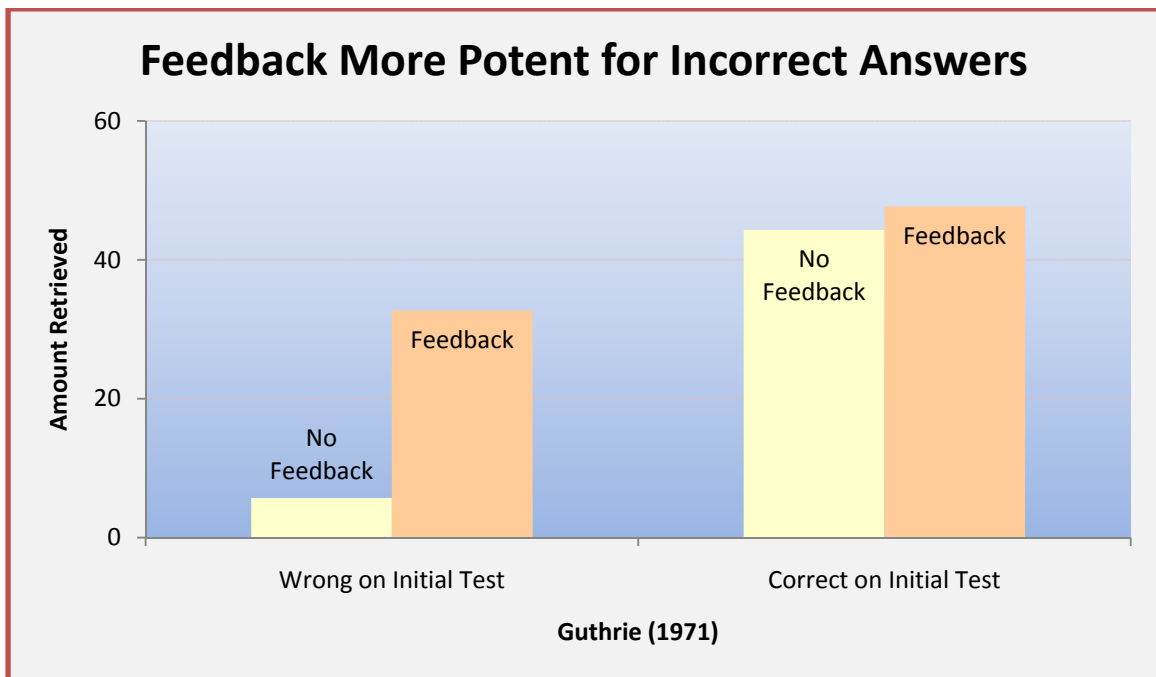
When no differentiation is made between feedback for correct and incorrect answers, benefits are still forthcoming, but often at less robust levels. Karraker (1967) found that feedback improved long-term memory performance over conditions that got no feedback by 10%. Kulhavy and Anderson (1972) found that feedback produced better results than no feedback by an average of 103%. Kulhavy, Yekovich, and Dyer (1976) found feedback was beneficial over no-feedback conditions by 15% when the retention interval was a week in length. Clariana, Ross, and Morrison (1991) found 26% improvement due

³ In addition to presearch availability, other major methodological weaknesses pollute the earlier literature on feedback. For example, most experiments only used immediate tests of retention and didn’t test retention after more than 24 hours. Thus, they may be measuring short-term effects that have no bearing on long-term real-world performance (see Schmidt & Bjork, 1992, for similar sentiments). Some experiments, notably list-learning experiments (like a large proportion of those reviewed by Kulik & Kulik, 1988), measure how quickly information is learned rather than retention after a delay. Thus, these experiments don’t even measure memory retention at all, let alone long-term retention. Similarly, the research on feedback often confuses different terminology. For example, telling a learner that their answer was wrong is considered feedback by some reviewers (Bangert-Drowns, Kulik, Kulik, and Morgan, 1991) even though it’s clear that feedback works because it corrects errors—since such right/wrong feedback does not correct errors it should not even qualify as feedback.

to feedback. Sassenrath & Gaverick (1965) found that feedback improved long-term performance by 12% in a naturalistic setting where learners were free to study on their own after the testing-feedback event. Webb, Stock, and McCarthy (1994) found improvements due to feedback in two separate experiments of 67% and 97%. Brosvic, Epstein, Dihoff, and Cook (2005) found an average of about 93% improvement due to feedback as compared with a condition that didn't get feedback.

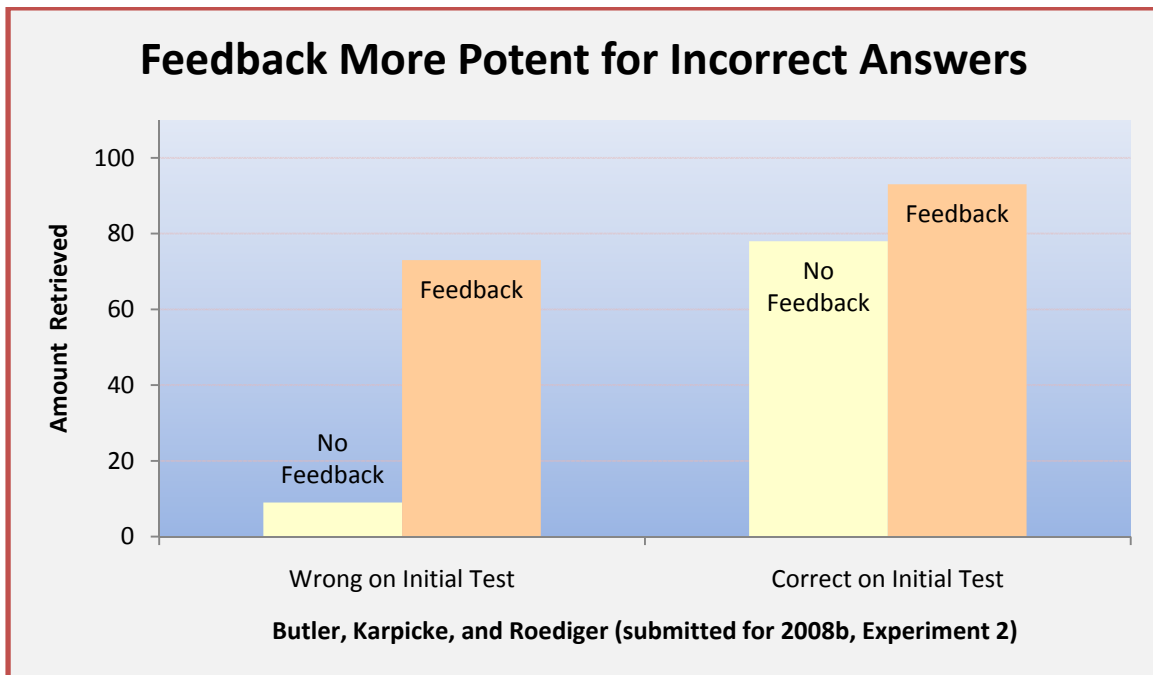
Feedback on Correct versus Incorrect Answers

As alluded to just above, feedback is considerably more important for incorrect answers than for correct answers. Feedback works by correcting errors. Guthrie (1971) found that feedback had almost no effect on correct answers (improving results by 8 percent over a no-feedback condition, a statistically non-significant difference), but produced major improvements when provided for incorrect answers (improving results by a whopping 474 percent!!).

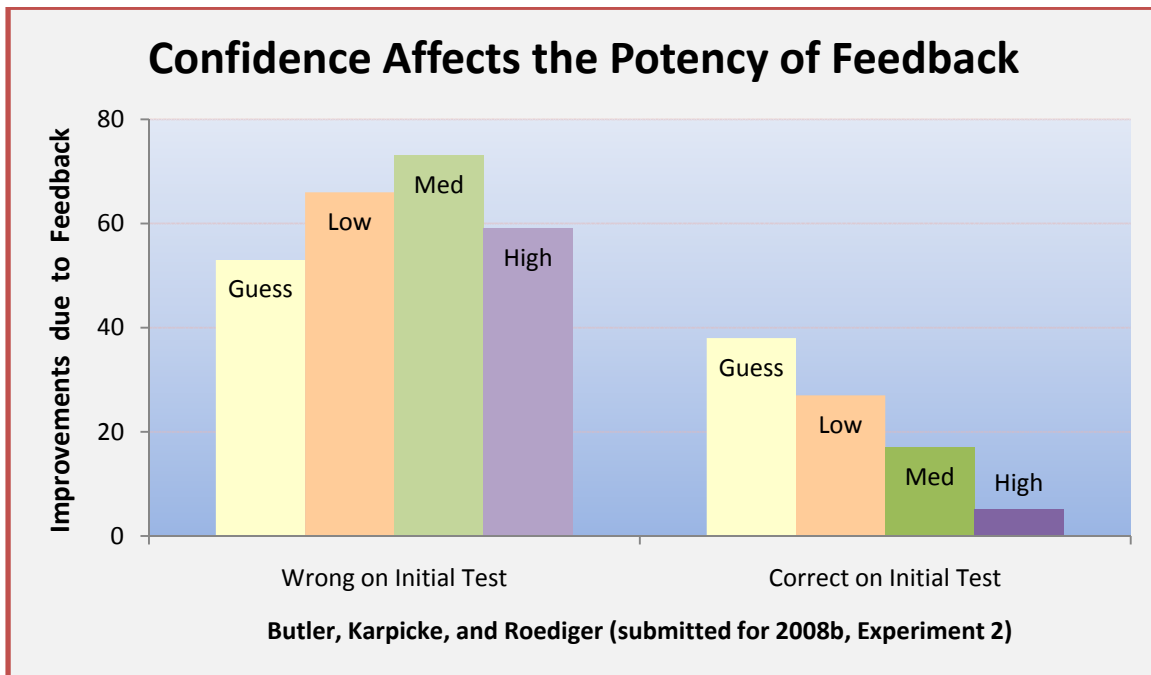


Surber and Anderson (1975) obtained similar results, finding that feedback had no effect on answers that were correct on the initial test, but had a significant effect on answers that were incorrect, improving later retrieval on those questions by an average of 47%.

But not all correct answers are the same, as is clearly shown in a study by Butler, Karpicke, and Roediger (submitted for 2008b). They asked learners how confident they were in their answers on an initial test. Their results for both these initially incorrect and initially correct answers are enlightening. The graph below shows the difference in the amount retrieved (after two days) when learners are given feedback compared with when they are not given feedback.



While the above graph shows the typical finding—that feedback on incorrect answers is much more important than feedback on correct answers—Butler, Karpicke, and Roediger added an ingenious twist to their experimental design. They asked learners how confident they were on the initial test (the one for which feedback might be given). The graph below shows the importance of feedback for each level of confidence (Experiment 2).



To examine the graph immediately above, note first that the leftmost set of four bars is significantly taller than the rightmost set of bars. If you add up the totals represented in the bars for incorrect answers⁴, you get 251 percentage points of change due to feedback. For correct answers⁵, it's only 87 total percentage points of change. Again, this demonstrates the greater importance of feedback for incorrect answers than for correct answers, but note that this graph actually underestimates the difference⁶. Note also that, contrary to some of the other research cited in the paragraphs just above, feedback created benefits for correct answers.

The leftmost set of four bars shows the power of feedback when learners get the answer wrong on an initial test. Feedback is very powerful to correct these wrong answers. The amount of confidence⁷ learners have in their initial answers doesn't show a clear trend. In

⁴ Guess = 53%, Low = 66%, Medium = 73%, High = 59%; totaling 251 percentage points. These were estimated based on graphical depictions presented in the research paper.

⁵ Guess = 38%, Low = 27%, Medium = 17%, High = 5%; totaling 87 percentage points. These were estimated based on graphical depictions presented in the research paper.

⁶ This graphic can't tell the whole story because the number of responses under each bar is not the same. For example, there are probably more "high-confidence" correct answers than "guess" correct answers, and more "guess" incorrect answers than "high-confidence" incorrect answers. The bottom line is that this graph underplays feedback's greater importance for incorrect answers. See the previous graph for a better representation of the relative importance of feedback for incorrect and correct answers.

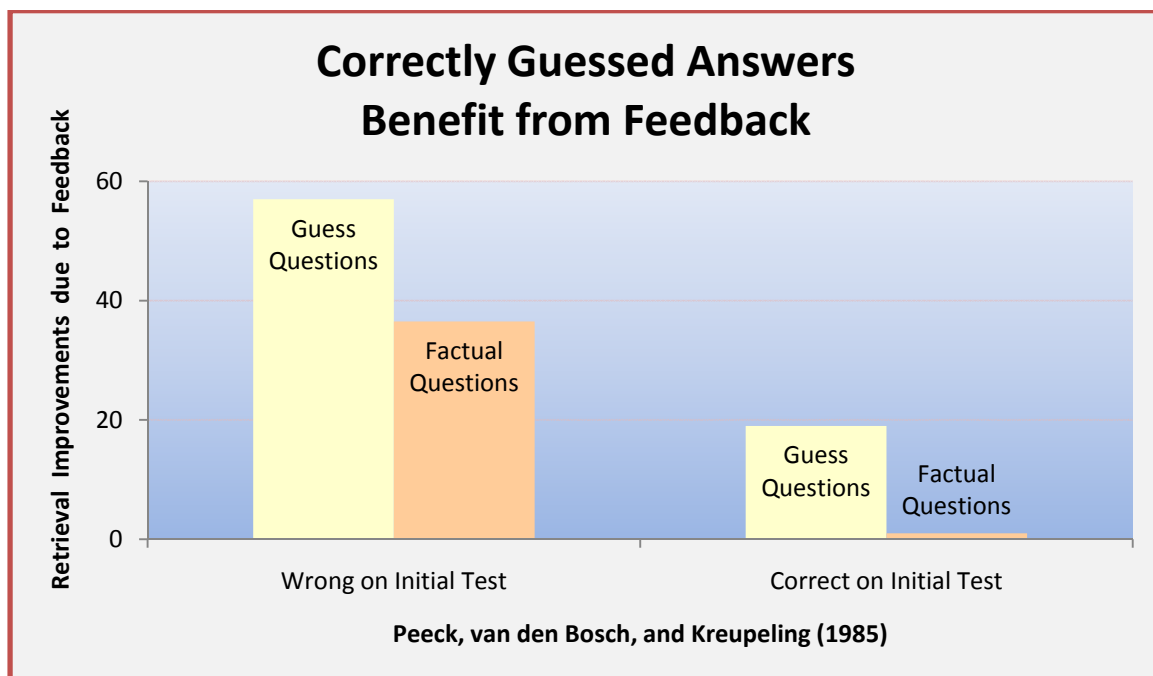
⁷ The levels of confidence displayed in the graph above (e.g., "Guess," "Low," etc.) are calculated by asking learners to rate their level of confidence that their answer will be correct on a numerical scale. For example, the "Guess" label represents those estimates that fall in the 0% to 25% range (the chance of being correct). "Low" is for low confidence or a 26% to 50% chance of being correct. "Med" is for medium confidence or a 51% to 75% chance. "High" is for high confidence or a rating of 76% to 100% chance of being correct.

other words, it isn't clear (for these initially incorrect answers) that the level of confidence will reliably affect the power of feedback to improve subsequent learning.

The rightmost set of bars shows that feedback does improve performance even when learners get answers correct on the initial test. In addition, confidence on those initial test responses shows a clear trend line. Lower-confidence responses improve the most with feedback. High-confidence initially-correct answers improve hardly at all. Of the 87 percentage points of change represented in the bars for correct answers, medium- and high-confidence answers comprised only 12 percentage points, a fairly miniscule amount.

A study by Peeck, van den Bosch, and Kreupeling (1985) didn't measure confidence, per se, but took it into account by forcing learners to guess on some questions and not others. The questions they guessed at (they had to guess because they weren't given any prior exposure to relevant learning material) showed significant improvements due to feedback, for both incorrect answers (380% improvement) and correct answers (28% improvement). This compares with the factual questions (the ones that they didn't have to guess at because they had spent time learning the material), which showed improvements due to feedback for incorrect answers (183% improvement) and correct answers (non-significant 1% change).

As can be seen on the graph of this experiment below, the "Guess Questions" benefited from feedback on correct answers, while the "Factual Questions" did not.



Again we see that correct answers are not all created equal. Some represent knowing, and some represent varying shades of confidence. When learners are guessing correctly (as opposed to knowing correctly), feedback improves their subsequent memory.

The Importance of the Correct-Incorrect Distinction

The distinction between correct and incorrect answers highlights one of the most important principles of feedback. Feedback works primarily by correcting errors—either detected errors (when learners respond incorrectly) or hidden errors (when learners respond correctly but guess at some varying level of confidence).

This framing highlights a key instructional recommendation—we need to uncover errors, help learners understand them, and then give learners additional practice to reinforce correct responding.

Feedback Enables Correct Retrieval Practice and Follow-up Retrieval Practice Produces Benefits

Although feedback helps correct errors, it can also prepare learners for future retrieval situations by enabling them to practice correct retrievals. Phye and Andre (1989) found that when learners re-responded to test questions after getting corrective feedback, they improved their later retrievals by 26% (when they received delayed feedback) and 97% (when they received immediate feedback) over conditions where they got feedback but didn't re-respond to the test questions. These results show the power of retrieval practice. Similar results have been found by Sturges (1972) who found that testing after feedback improved long-term retention on a recall test by about 95% and on a recognition test by about 20%. That recall tests are more susceptible to retrieval practice is expected because recall tests require unassisted retrieval whereas recognition tests provide alternatives that enable learners to assess the cognitive connection between the question and each alternative.

Of course, in both of these examples, it was the feedback that enabled learners to practice the correct response. Without feedback, retrieval practice might solidify inappropriate knowledge. Evidence that retrieval practice without feedback is completely ineffective was supplied by Kulhavy and Anderson (1972) who gave learners as many as three tests without feedback and compared them with learners who got one test and feedback. The learners who got feedback and only one test surpassed those who got only three tests by an average of 103%.

Note that this analysis highlights once again that feedback doesn't work in isolation. Before learners get feedback, they've already practiced retrieving information. If they've retrieved information correctly, their knowledge will tend to remain intact and retrievable through the feedback phase to a later performance test. However, if learner's initial retrieval practice is incorrect, feedback can help them correct their errors. But it's even better to provide learners with feedback and than a chance to re-respond to those initial queries. This additional retrieval practice increases the likelihood that the learner will overcome the effects of their initial incorrect retrieval practice. Again, retrieval practice is the key.

The Peeck, van den Bosch, and Keupeling (1985) result is again instructive. Recall that they had learners guess blindly on an initial test (for half the questions), receive feedback or not, and then take a final test. The learners scored at chance probability on the initial test when guessing, scoring only 27% correct. However, what's remarkable is that even though the learners were guessing blindly on the initial test, the answers they got correct on that test had a strong chance of being correctly answered on the final test. With feedback, these correct guesses got turned into 86% correct responding on the final test. Okay, that just shows the power of feedback. But even when they got no feedback, the correct answers from the initial test were reproduced correctly on the final test at a whopping 67% rate. Again, the initial retrieval opportunity—even without feedback—produced very strong results on its own.

Should Learners Get Feedback Only After Making a Response?

In the 1960s, a whole slew of research studies found that providing feedback actually hurt learning. These results caused several learning theorists to jump off buildings⁸ and pushed the whole profession into a chaotic fury. Fortunately, these results were explained by Anderson, Kulhavy, and Andre (1971, 1972), who discovered that when learners could “peek” at feedback before fully responding to questions, they would peek. In fact, it appeared that learners couldn't help themselves.

More terrifying was the result. Learners who got feedback did worse than those who got no feedback because those who got feedback hadn't fully processed the initial test questions before peeking at the answers. Based on the research of Anderson, Kulhavy, and Andre (1971, 1972), and later commentary (Kulhavy, 1977), feedback was returned to its exalted state and the importance of having learners answer questions before getting feedback reverberated throughout the research world.

Notice what was happening. Learners who peeked at the answers did worse on subsequent tests of retrieval because the act of forgoing the retrieval opportunity (by peeking instead of retrieving) had harmed their learning. Again, we see the power and importance of retrieval.

The bottom line is that when we're preparing our learners to retrieve, we need to make sure that our feedback mechanisms don't short-circuit retrieval practice.

On the other hand, the Anderson, Kulhavy, and Andre (1971, 1972) research has very little to say about providing feedback to help people understand (as opposed to supporting future retrieval). The learning material they used was programmed instruction from the behaviorist tradition where learners basically had to remember very simple

⁸ I'm joking. No lives were lost as far as I know.

concepts they had just read. Feedback in these experiments, therefore, was designed to support retrieval, not to facilitate understanding per se.

So, the jury is still out when using feedback to help learners understand learning concepts. Given the current state of the research, I don't see any reason why you shouldn't experiment by letting your learners—early in the learning process—peek ahead to feedback as they are trying to understand and create mental models of complex learning materials. However, once the learners “get” the concepts, don't give them feedback until they've completed their retrieval-practice attempt.

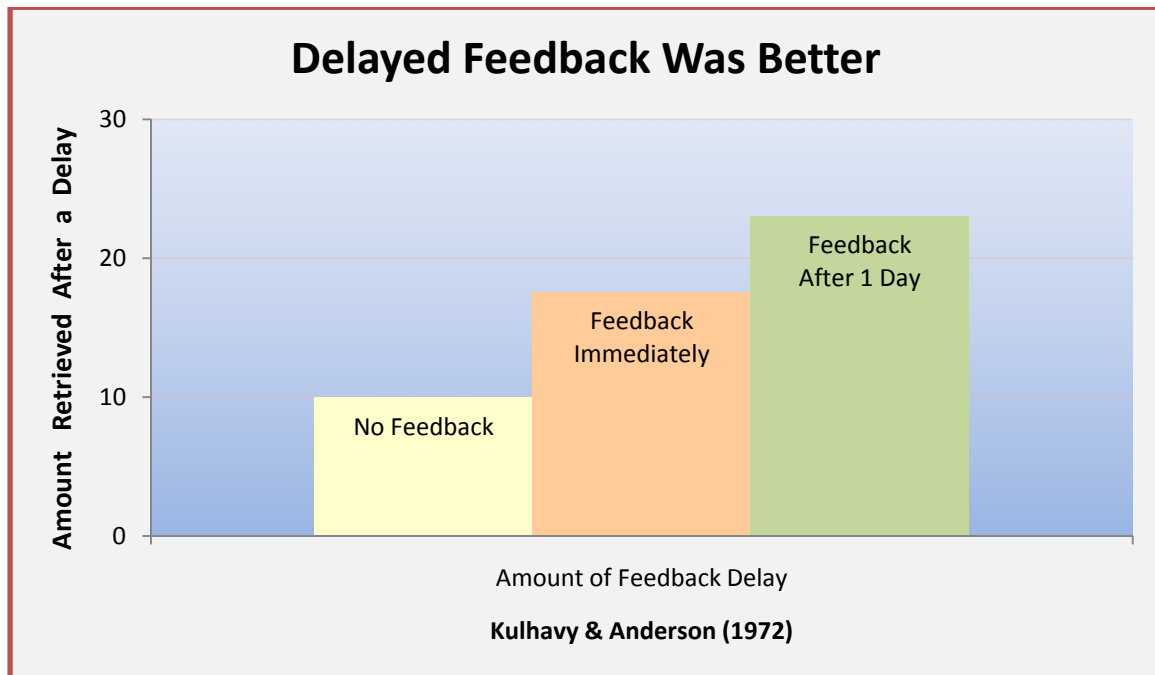
Timing of Feedback—Which is Better, Immediate or Delayed Feedback?

When should feedback be given, immediately or after a delay? A tremendous amount of ink has been spilled in attempts to answer this question, but a lot of what has been communicated is smeared and blotched in confusion, and remains so, as we will see.

It was once believed that immediate feedback was always better than delayed feedback. This vestigial belief was borne of behaviorist notions of reinforcement (e.g., Skinner, 1954; 1968) but was debunked starting in the 1960s and 1970s with a series of findings that delayed feedback was better than immediate feedback (Sassenrath & Yonge, 1968, 1969; English & Kinzer, 1966; More, 1969; Sturges, 1969, 1972; Phye & Andre, 1989).

Sassenrath and Yonge (1968) gave learners a test immediately after giving them feedback and then a second test 24 hours later. On the immediate test, learners who got delayed feedback performed the same as learners who got immediate feedback. However, on the more important and realistic delayed test, learners who got delayed feedback performed better than learners who got immediate feedback.

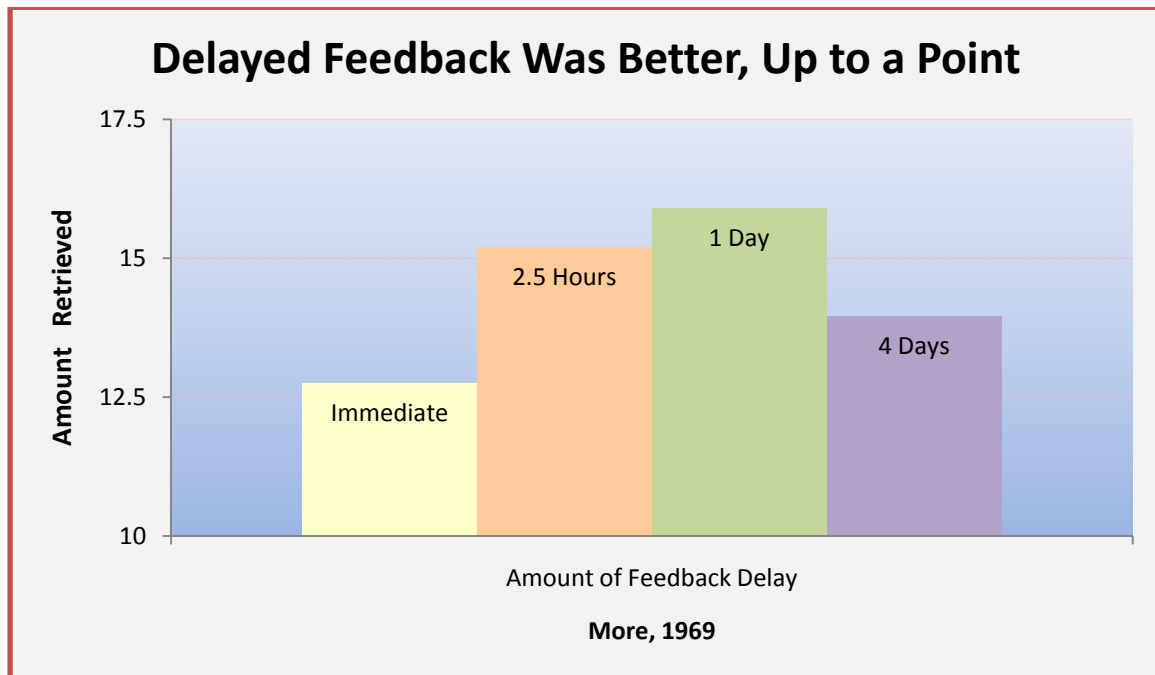
This pattern—that delayed feedback is particularly potent for long-term retention—was found in a number of studies at the time (Kulhavy & Anderson, 1972; More, 1969; Sturges, 1978).



To reiterate, these studies showed that feedback is most important for long-term retention. In fact, one of the reasons that much of the previous research showing advantages for immediate feedback was flawed—and thus not very useful—is because the researchers used only immediate tests of retention. When tests are given immediately after learning, they are poor predictors of long-term retrieval. Of course it's important to remember that most training situations aim for long-term retention, not simply immediate retention, so long-term retention is a much more relevant metric.

Note that delayed feedback, while found mostly to benefit long-term retention, has also occasionally been found to benefit short-term retention as well. For example, when Sassenrath (1975) reanalyzed the data from Sassenrath and Yonge (1968), he found that initially incorrect questions showed more of a benefit from delayed than immediate feedback on both an immediate test (by 30%) as well as on a delayed test of retention (by 40%). Thus, we cannot completely rule out the benefits of delayed feedback for immediate retention, because it may have some special effect on initially incorrect answers.

How long feedback should be delayed is not clear. Many of these studies showed an advantage to delaying feedback for a day. Other studies have found an advantage for delaying feedback for 20 minutes (Sturges, 1978), one hour (English & Kinzer, 1966), and 2.5 hours (More, 1969). One study found that delaying feedback for four days actually produced less learning than delaying feedback for one day (More, 1969), so it appears that feedback may lose its power if it is delayed too long.



The optimal delay probably depends on a number of factors, including the type of material to be learned and the amount of time before the information will be retrieved, but researchers haven't clearly worked out what delays work with what factors.

Research Reviews on Feedback Timing are Not Definitive

To emphasize the lack of clarity in the research related to the timing of feedback, research reviews have come to different conclusions. Kulik and Kulik⁹ (1988), Kulhavy, 1977; Bangert-Drowns, Kulik, Kulik, and Morgan (1991), Mory (2004), and Azevedo and Bernard (1995) concluded that immediate feedback is better, whereas Kulhavy and Stock (1989) concluded that delayed was better, and still others discuss the pluses and minuses of feedback timing (Renner, 1964). Unfortunately, the comparison isn't fair

⁹ The Kulik and Kulik (1988) research review has often been quoted as definitive evidence for the benefits of immediate feedback. Unfortunately, the research studies it reviews are plagued with problems. First, many of the studies reviewed use immediate tests of memory. Again, such tests would not be expected to show benefits for delayed feedback because delayed feedback creates its benefits through a spacing effect, which shows benefits primarily for long-term retention. The second major methodological problem in the studies reviewed by Kulik and Kulik (1988) is the use—as a dependent measure—of the number of trials till all the pairs of a list were acquired. This is a particularly dangerous design because it doesn't measure long-term retention at all. Moreover, this design's pseudo measurement—the number of trials till acquisition—is probably a causal factor itself in producing long-term retention. In other words, because more trials till acquisition provides the learner with more retrieval practice on those pairs, what seems like a bad thing (it takes more trials until the all information is learned) might actually be a good thing (more retrieval practice enables better long-term retention). The bottom line is that this design is simply too likely to lead to confusion and experimental confluations. Kulik and Kulik's (1988) conclusions must be discounted due to these substantial methodological weaknesses in the research they reviewed.

because all the pro-immediate-feedback reviews fail to consider the importance of the length of the retention interval. This is a major oversight because delayed feedback appears to work in a similar manner to the spacing effect, which has been clearly shown to create benefits for long-term retention intervals, and much less often for short retention intervals (for example, for tests given immediately after learning). For reviews see, Cepeda, Pashler, Vul, Wixted, & Rohrer (2006); Donovan and Radosevich (1999); Lee and Genovese (1988); Ruch (1928); Cain and Willey (1939); Melton (1970); Crowder (1976); Hintzman (1974); Glenberg (1979); Rea and Modigliani (1988); Dempster (1988, 1989; 1996); and Thalheimer (2006).

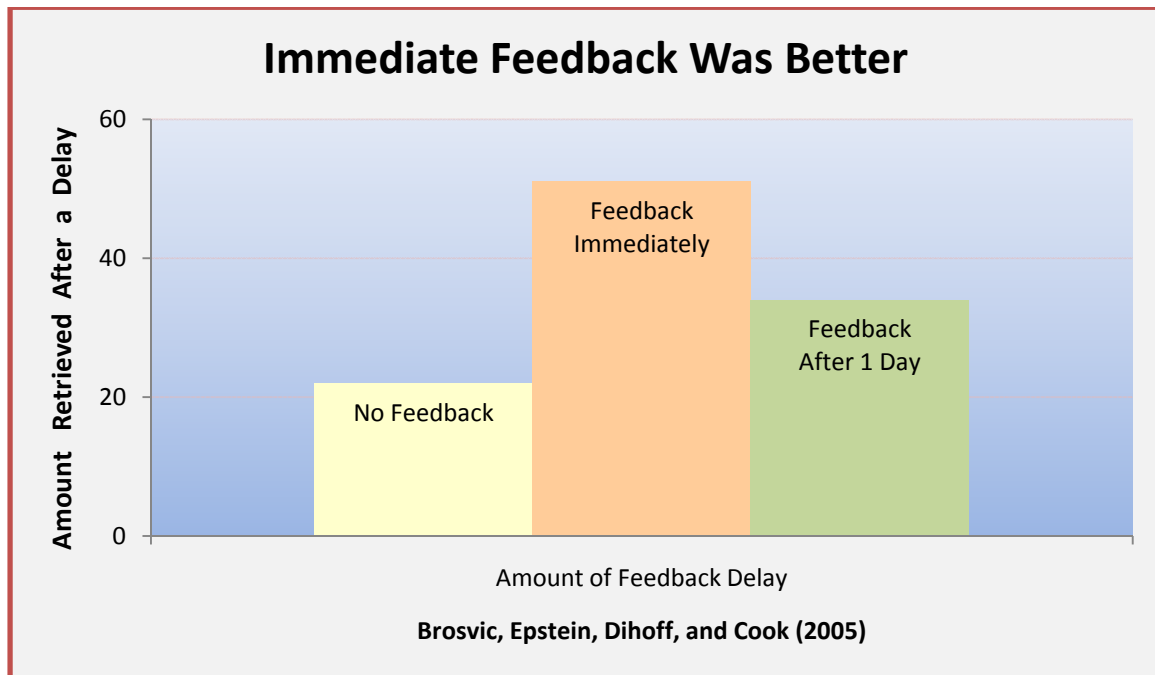
In other words, whereas for short retention intervals we shouldn't expect advantages for delayed feedback over immediate feedback, we can expect advantages due to delayed feedback's spacing benefits for more realistic long-term retention. Therefore, the experiments that used short retention intervals are, for all practical purposes, irrelevant to the question of immediate versus delayed feedback. Moreover, the reviews based on these experiments must be discounted as well.

Two New Research Threads on the Timing of Feedback

Recently, two new research threads on feedback have risen to prominence, again muddying the waters about which is better, immediate or delayed feedback.

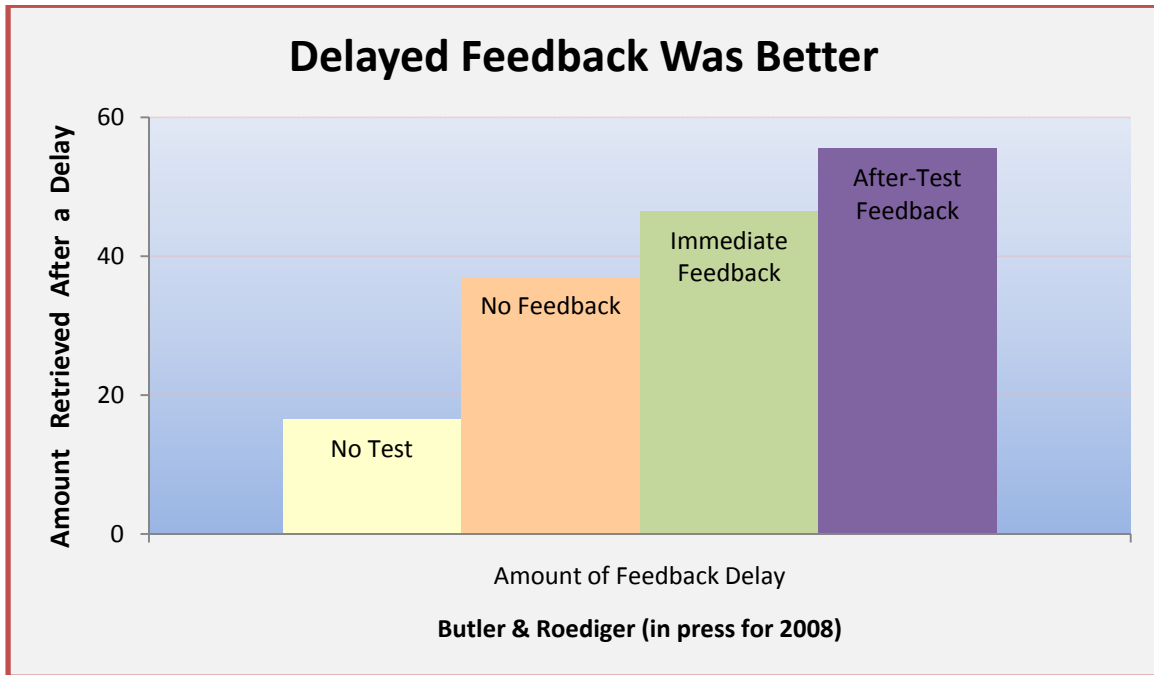
Using a scratch-off answer sheet (called the IF AT, the Immediate Feedback Assessment Technique form) that enables immediate feedback for learners taking tests, Brosvic, Epstein, Dihoff, and Cook (2005) found benefits to immediate feedback over delayed feedback in a fairly realistic college classroom environment. Improvements of 50% for immediate after-answer feedback over feedback delayed either until the end of the test or until the following day. Similarly, Brosvic, Epstein, Dihoff, and Cook (2006), using laboratory experiments designed to simulate classroom learning, found improvements of 9% for immediate after-answer feedback over feedback delayed either until the end of the test or until the following day.

The results of Brosvic, Epstein, Dihoff, and Cook (2005) are displayed visually in the following graph:

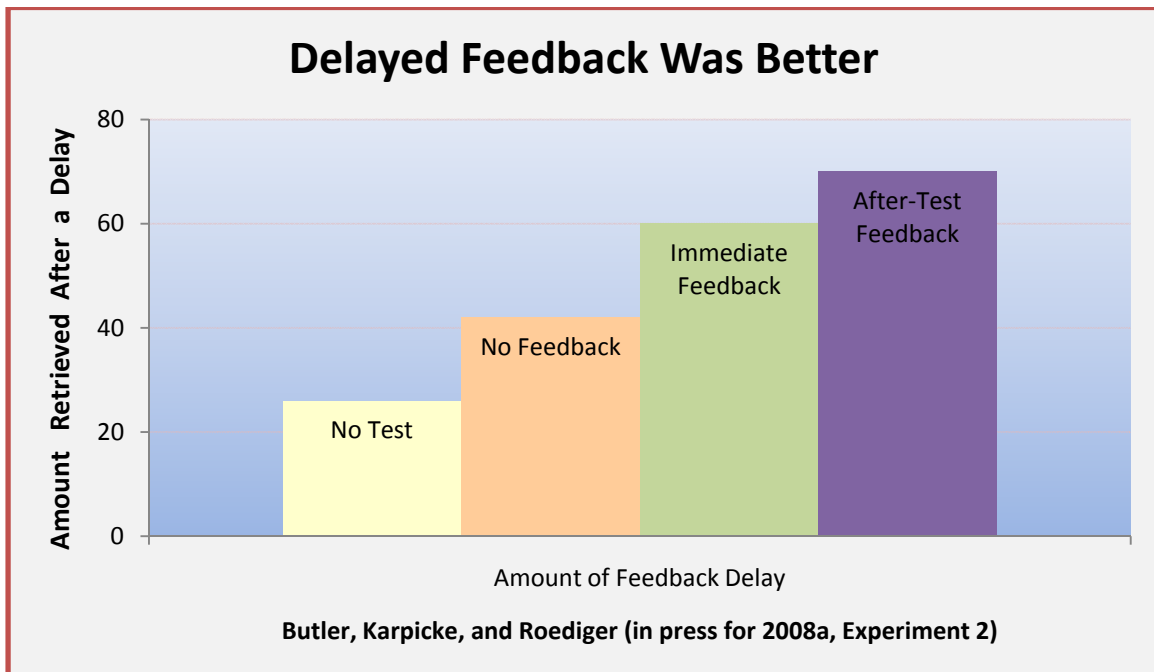


Unfortunately, this line of research suffers (a) from confusing experimental write-ups, (b) potential bias due to at least one researcher’s commercial interest in the IF AT product line, (c) a research history (prior to the studies cited above) peppered with conflations and methodological weaknesses, and (d) article publications that have been restricted to second-tier research journals. Despite these indications, it may not be fair to ignore the research results out of hand. The research paradigm differs enough from other lines of research that it is possible that the findings may be due to learning factors, not to sloppy or biased research methods. Specifically, this line of research differs from many other research paradigms in using realistically comprehensive learning materials (as opposed to simple informational items) and enabling learners to study between the presentation of feedback and the subsequent final retention test. Until other researchers attempt to fully replicate these results, it would be premature to ignore the aforementioned findings.

The second new line of research—methodologically on firmer footing—finds the opposite result, that delayed feedback is superior to immediate feedback. Butler and Roediger (in press for 2008) found that after-test feedback was better than feedback given after each question by 19% on a multiple-choice final test and 21% on a free recall final test. See below for the multiple-choice results:



Note also the leftmost “No Test” bar, which in comparison to the “No Feedback” condition reiterates the power of retrieval practice itself (that is, because “No Feedback” implies that the learners got a test). Also notice that feedback, whether given immediately after each test question or delayed until after the whole test, produced better learning outcomes than no feedback. Similar benefits to delaying feedback have been found in Butler, Karpicke, and Roediger (in press for 2008a). Results from their Experiment 2 are detailed below, showing a 17% improvement in retrieval due to delayed feedback.



How to Make Sense of the Muddled Research Results on Feedback Timing

First, it's important to acknowledge that both immediate and delayed feedback produce significantly better learning results than conditions that provide no feedback. In fact, even researchers who recommend delayed feedback (for example, Butler, Karpicke, & Roediger, in press for 2008a) have subsequently used immediate feedback in their experiments (Butler, Karpicke, & Roediger, submitted for 2008b). Moreover, many experiments find benefits of immediate feedback when using the simple comparison of immediate feedback to no feedback (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005).

At the present time, the research advocating the use of delayed feedback seems more credible overall than the feedback advocating for the use of immediate feedback. However, there are enough discrepancies and unconsidered factors to make such a direct comparison a dubious enterprise. From a research perspective, we must wait for additional results that tease apart these many factors. From a practical perspective, we have to recommend some basic findings in regard to research timing, including (a) the benefits of feedback, whether it is immediate or delayed, (b) the greater importance of feedback for incorrect answers, and (c) the importance of the attention given feedback (which may depend on feedback timing for different learners).

The Distinction Between Knowledge, Marginal Knowledge, and Submarginal Knowledge.

Harry Bahrack and colleagues (Bahrack & Hall, 1991; Bahrack & Phelps, 1988; Berger, Hall, & Bahrack, 1999) have provided an important model regarding the ease of memory retrieval that offers significant practical value. They compare information that can be retrieved with some form of memory aid (like multiple-choice alternatives) to information that can be recalled without such aids. Berger, Hall, and Bahrack (1999) define information that can be retrieved only with recognition aids like multiple-choice alternatives as “marginal knowledge.” They label information that cannot be recalled or recognized as “submarginal knowledge,” perhaps alluding to the reality that information, once learned, is never truly lost from memory—but rather it is in memory but is simply not accessible. To take their labeling further, we might call information that can be retrieved without special memory aids as “knowledge.”

As they suggest (Berger, Hall, & Bahrack, 1999, p. 438), “the practical value of marginal knowledge is limited.” Information that can only be recognized in the presence of specific identifying cues is not very valuable because most information must be recalled without such powerful hints.

They take this notion further by suggesting that one of the goals of learning design should be to ensure that marginal knowledge is made more retrievable. They demonstrated this in their research by showing how information that could not be recalled initially could be recalled later if it was corrected through an intervention in the interim that (1) prompted

the learners to answer a follow-up question that included both the question stem plus the addition of five multiple-choice answers (making correct retrieval easier than with the question alone) and then (2) followed up the multiple-choice question with feedback that told the correct response. This two-pronged “feedback” mechanism increased the amount retrieved after one day by 580%, and after nine days by 308% over control conditions that provided no feedback. These improvements are huge because the baseline comparison is represented by information that learners couldn’t recall initially. So in some sense, they compared a learning condition (the feedback condition) to a no-learning condition (the baseline no-recall condition).

These results emphasize again the importance of retrieval and feedback. They also show us how forced-choice recognition questions can facilitate later recall, and in some ways may act as feedback themselves. Finally, the results point to the importance of helping learners minimize the effects of forgetting by providing subsequent and routine learning opportunities, of which feedback and retrieval practice can be key components.

Feedback’s Effects Can Be Minimized

Feedback’s effect can be stunted under some conditions. Feedback seems to have less of an effect in natural settings where learners can spend time studying after they get or don’t get feedback. For example, Moore and Smith (1961) found no differences between feedback and no-feedback when they tested the effects of feedback over five weeks in a real classroom setting. Newman, Williams, and Hiller (1974) found no differences in a normal classroom setting over a seven-day period. These results are not unexpected, nor should they be seen as evidence against the effects of feedback. In these situations, learners who don’t get intentionally-didactic feedback after a test are likely to restudy the material and provide their own feedback by processing information relevant to the questions they were asked. Such opportunities are likely to put those with no feedback on par with those who received feedback. Despite this likelihood, some researchers have found that even under these difficult conditions—where learners can restudy after the feedback situation—feedback can produce better results than no-feedback conditions (Sassenrath & Gaverick, 1965; Brosvic, Epstein, Dihoff, & Rohrer, 2005).

Feedback for Physical Movement Skills

Most of the research on feedback is done in regard to declarative knowledge, not physical movement skills. Declarative knowledge is knowledge about concepts—about “what is.” It differs from procedural knowledge which is about “what to do.” Because the research on feedback is heavily weighted toward knowing, not doing—and because the ultimate goal of most meaningful learning is preparing people for doing, and because physical movement is all about doing—it’s my hunch that the research on physical movement skills may give us a window into feedback that the bulk of research on declarative knowledge is not yet capable of providing. So, for example, where the physical

movement research finds that more extensive feedback is particularly beneficial for beginners, it seems likely that for complex procedural knowledge, extensive feedback may be beneficial for beginners as well. Research on simple declarative knowledge is simply not relevant to this question.

My review of the physical-movement skill literature is drawn primarily from two extensive research reviews (Magill, 1993; Salmoni, Schmidt, & Walter, 1984) with augmentation from additional and more recent research articles.

A small number of research studies are consistent with the notion that feedback on physical movements works best when it corrects errors rather than when it attempts to reinforce correct movements (Magill, 1993). This is consistent with the research on feedback for cognitive learning tasks we discussed earlier in this report (e.g., Guthrie, 1971). However, the notion that feedback works by correcting errors wasn't even discussed in the review on movement learning by Salmoni, Schmidt, and Walter (1984), an indication that relatively few studies have examined this issue. Because only a few studies have directly explored this question, more research is needed before we can be certain that feedback in movement learning creates its effects only by correcting errors.

Several lines of research described in Magill (1993) lead me to conclude that beginners benefit from significant amounts of feedback early in the process of learning movement skills. Research has found that biofeedback can be used to help people learn to monitor their heartbeats, a skill for which most people are beginners. Redundant feedback, although harmful for most learners, is beneficial for beginners because it enables them to get a feel for how their new movements are creating their effects. Recent evidence that complex skills may require more feedback rather than less feedback (Wulf, Shea, & Matschiner, 1998) may also be relevant here. When beginners face new learning situations, they usually confront complexities they haven't dealt with before. Thus, by extension, if complex tasks benefit from more feedback, then beginners dealing with new complexities are likely to benefit from more feedback as well. Obviously, the research conclusion that beginners benefit from feedback early in their learning is somewhat tenuous.

Until recently, most researchers felt that most movement learning tasks benefited from providing less feedback rather than more feedback. However, at least one recent study has found that the most complex skills may benefit from more feedback, at least initially in the learning process (Wulf, Shea, & Matschiner, 1998). Because this is a relatively new line of research, much more research is needed before any conclusions are definitive.

A large number of studies and reviews of research (Magill, 1993; Salmoni, Schmidt, & Walter, 1984) argue that feedback that is redundant with environmental stimuli should be used sparingly, or intermittently, or not at all. Feedback is viewed as providing guidance during practice (Salmoni, Schmidt, & Walter, 1984). As mentioned above, such guidance can prompt learners to ignore the stimuli they'll need to monitor when they are faced

with their real-world performance situations. The research on this redundancy effect is strong and was found in a large number of experimental situations.

Several lines of research converge in recommending that feedback should be given periodically instead of being provided on every practice trial. As Magill (1993) reports, giving feedback on less than 100% of practice trials is almost always superior or equal to giving performance on all the practice trials. Providing summary feedback after a series of practice trials is better than providing feedback on every trial. Allowing learners to perform several trials before providing feedback on each of those trials is generally more beneficial than providing feedback after each trial.

Giving learners feedback only when their performance falls outside of certain performance bandwidths is consistent with a number of research studies (Magill, 1993). Note that it is unclear whether this effect is due to the bandwidth mechanism or whether it is due simply to the fact that feedback is not given on every practice trial. More research could help clarify this and provide practitioners with more specific guidance in deciding between different feedback practices.

Researchers have extensively investigated the effects of different learner activities after feedback is given. Research shows that when learners use the time they have between feedback and the next practice trial to evaluate their performance; they perform better in the long run (e.g., Swinnen, 1990). Based on this research, Magill (1993, p. 208) recommends that learners be asked, “‘What do you think you did wrong?’ before being told what they did wrong.” Although there are only a small number of studies on this particular point, the ones that have been done consistently show benefits from this type of learner self-evaluation. I might add to this notion. It might also be worth asking learners first, “‘How do you think you did?’” This encourages them to monitor their own performance. Asking them what they did wrong may prompt them to wait for feedback instead of engaging in monitoring activities.

Several recent studies suggest that feedback should not be given immediately after a practice trial, but should be delayed at least for several seconds (Anderson, Magill, Seklya, 2001; Swinnen, Schmidt, Nicholson, & Shapiro, 1990; Magill, 1993).

A large body of research has found that it is usually better to give people feedback after they’ve completed a task instead of giving them feedback while they are doing a task (reviewed by Armstrong, 1970, as reported by Magill, 1993).

There is a strong research base that shows that quantitative feedback is better than qualitative feedback except when learners are just beginning to learn something. More specifically, it has been found that the more precise the feedback the better, up to the point where learners can’t perceive the finer distinctions (Salmoni, Schmidt, & Walter, 1984; Magill, 1993).

Research has shown for some time that videotape feedback can be helpful for intermediate and advanced learners if the visual feedback in the video helps people correct their performances. Beginners require instructors to annotate the discussions (reviewed by Rothstein & Arnold, 1976, as reported by both Magill, 1993 and Salmoni, Schmidt, & Walter, 1984).

Although research has shown that excessive feedback can hurt long-term performance, there are at least a few studies (Lavery, 1962, 1964, reported by Salmoni, Schmidt, & Walter, 1984, p. 375) that suggest that instructions to learners to expect a real-world task without feedback may help them overcome the negative effects of redundant or excessive feedback. More research is needed before we can assume that such instructions will provide a quick fix to poor feedback designs.

Other Issues to Consider for Feedback Design

The Specificity of Feedback

Learners can be given feedback after answering questions, making decisions, or taking actions in many ways. In general, feedback has been found to be most effective when it provides specific guidance about what is correct. For example, corrective feedback improved long-term memory performance over conditions that got only feedback regarding whether answers were right and wrong by 61% with an effect size of 3.20 (Phye, 1991) and by 27% with an effect size of 1.33 (Phye & Sanders, 1994). Pashler, Cepeda, Wixted, and Rohrer (2005), examining the learning of foreign vocabulary words, found that providing learners with correct-answer feedback produced better results than providing them with correct-incorrect feedback.

Sassenrath and Gaverick (1965) gave learners a final test with questions they hadn't seen before on earlier tests. They found that providing feedback by providing the correct answer and allowing learners to ask additional questions in a classroom setting improved their final test performance by 7% (with an effect size of 0.39) over a condition where they received no specific feedback. Although this result is small, it was found in a naturalistic setting where the learners could study on their own—a situation that might normally negate the effects of feedback (because studying provides its own feedback).

Getting Learners to Pay Attention to the Feedback

Feedback isn't worth a damn unless learners pay attention to it. So for example, Brosvic, Dihoff, Epstein, and Cook (2006) and Dihoff, Brosvic, Epstein, and Cook (2005) found that immediate feedback provided to low-ability students by an educator was better than immediate feedback provided by an answer sheet alone. Their result suggests that these particular students benefitted from the personal nature of the feedback situation. Such benefits are not always found. For example, the same set of researchers using the same educator-delivered feedback did not find similar improvements with a different set of students (Brosvic, Epstein, Cook, & Dihoff, 2005).

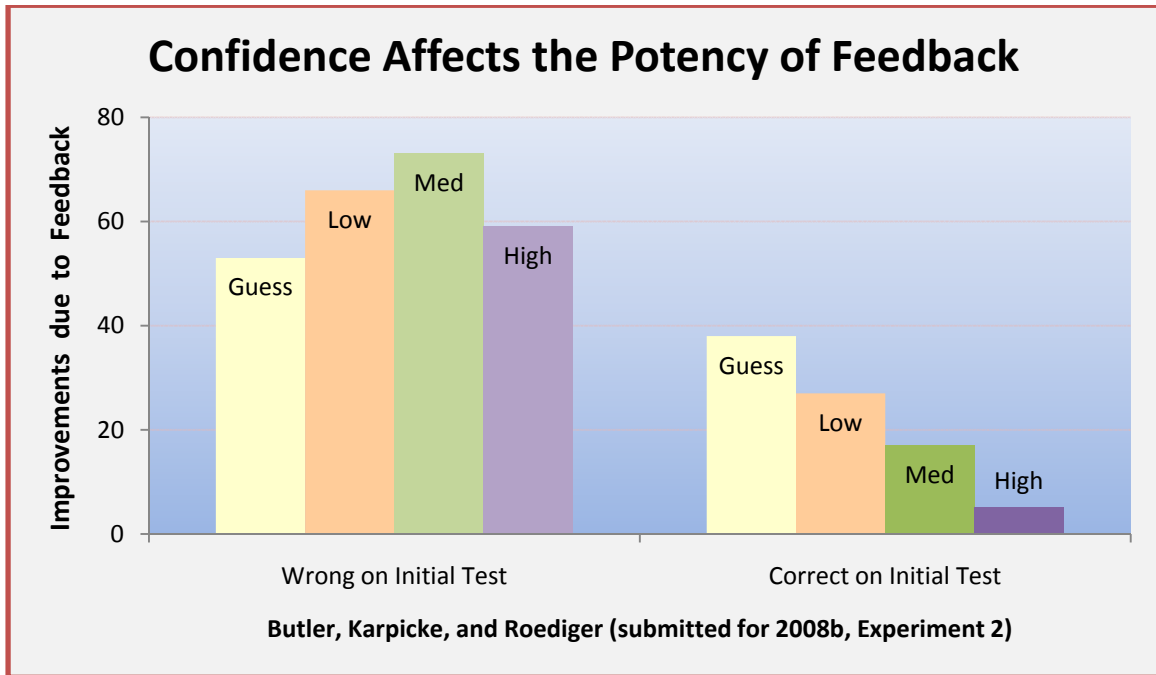
In the medical journal *Lancet*, researchers Thomas, Croal, Ramsay, Eccles, and Grimshaw (2006) reported that feedback provided to doctors to encourage them to use fewer inappropriate lab tests produced significant benefits. The feedback consisted of graphed data (sent to physicians four times in the one-year study period) comparing each practice's test-request rate to other practices in their region, plus short educational messages relevant to the specific type of lab test (for example, lab test "CA125 should not be used to screen, diagnose, or exclude malignancy."). Whereas a control group that received no feedback increased the number of tests requested during the one-year study period by 14%, the medical practices that received the feedback decreased their results by 12%, bettering the control group by 26%. This result may have more to say about learning materials than feedback, per se. On the other hand, providing busy doctors with brief memos four times a year is unlikely to have created as much of an impact if the doctor's own results hadn't been an integral part of the feedback.

Unfortunately, most experimental situations have demand characteristics that ensure that learners pay attention to feedback. The bottom line is that research-based recommendations may or may not produce effects in real-world situations—situations that don't prompt learners to exhibit the same level of attention to the feedback. As always, practitioners will need to experiment with what works for their learners in their situations.

Using Confidence Ratings to Augment Feedback

Confidence is clearly related to learners' ability to retrieve information. The more confident learners are in their ability to retrieve a particular piece of information, the more likely they will be able to retrieve that information in the future (Pashler, Cepeda, Wixted, & Rohrer, 2005; Butler, Karpicke, & Roediger, submitted for 2008b). Given this reality, we might wonder whether confidence ratings might be used to provide learners with feedback. Shouldn't it be obvious that learners with lower confidence on a particular learning point might require more or better feedback than those who are confident? Several commercial applications are now available that purport to find benefits from such confidence-based feedback approaches.

While the research on this point is still fairly thin, recent research by Butler, Karpicke, and Roediger (submitted for 2008b) is instructive. They found that low-confidence correct answers were more susceptible to feedback-induced improvements than higher confidence correct answers. Moreover, they found that while feedback was critically important on incorrect answers, confidence ratings did not provide any additional improvements. In the graph below, note how for incorrect answers (shown in the leftmost bars), confidence ratings did not follow a reliable trend, whereas for correct answers (the rightmost bars) the trend is clear—the lower the confidence, the more potent the improvement due to feedback.



Other research on confidence ratings and feedback have found similar lack of value in confidence ratings on incorrect answers (Pashler, Cepeda, Wixted, & Rohrer, 2005), except for the research that used unrealistically short retention intervals (for example, the five-minute interval used in Butterfield & Metcalfe, 2001).

Research that measured the amount of time learners spent reading paragraphs of feedback has generally shown that learners spend the least time processing feedback related to high-confident correct answers and the most time processing high-confidence errors (Kulhavy, 1977), lending credence to the idea that confidence mediates behavior related to feedback. In fact, Kulhavy and Stock’s (1989) theory of learner feedback processing is based on the notion of response certitude—another name for confidence.

Using the Answer-Until-Correct Methodology

While most forced-choice testing (for example, multiple-choice testing) requires test takers to select one answer and move on to the next question, the answer-until-correct method forces learners to select answer choices until the correct answer is chosen. This method can provide learners with greater credit when they utilize fewer guesses, (for example, learners can get full credit for correct first choices, 75% credit for second choices, 50% credit for third choices, and so forth). The answer-until-correct methodology can also be used with no grading or with full credit for a correct answer regardless of position.

Research cannot yet make a clear recommendation to help us decide between answer-until-correct feedback and standard feedback.

In a recent study, Butler, Karpicke, and Roediger (in press for 2008b), found that answer-until-correct feedback produced similar results to standard feedback, both when it was provided immediately and when it was delayed (for 10 minutes in Experiment 1 and one day in Experiment 2). They further found that second-choice answers produced better final retention scores than third, fourth, fifth, or sixth answers, presumably because second-choice answers were more established in memory.

Other researchers have found advantages for answer-until-correct feedback, including Brosvic, Epstein, Dihoff, & Rohrer (2005, 2006); Sullivan, Schultz, and Baker (1971); and Pressey (1950). While some of these studies suffer from methodological weaknesses—for example, conflating the type and timing of feedback (as noted by Butler, Karpicke, and Roediger, submitted for 2008b)—and may cause some of the multiple-choice incorrect lures to be better remembered; more research is needed.

To illustrate this further, note how in the Butler, Karpicke, and Roediger (in press for 2008a) article—which finds no difference between the answer-until-correct feedback and standard feedback conditions—learners who answered correctly on the second answer-until-correct option outscored folks getting the standard feedback. On the basis of this, we might conclude that the difficulty of the material (or the competence of the learners) may affect the comparative benefits of answer-until-correct and standard feedback. Maybe well-learned material (where learners are likely to get answers correct in one or two chances) would benefit more from answer-until-correct feedback, while less-well-learned material would benefit more from standard feedback. Perhaps other factors are predominant. For example, maybe the level of after-feedback interference is more important. In the Butler, Karpicke, and Roediger (in press for 2008a) experiments, learners experience relatively little in-context interference, whereas the Brosvic, Epstein, Dihoff, & Rohrer (2005; 2006) studies promote a significant amount of in-context interference, using final retention tests composed of 50% new items.

Bottom line is that it is still too early to make a call between answer-until-correct and standard feedback, though practitioners should feel free to experiment with both.

Note, however, that I have seen answer-until-correct feedback earn the impassioned disdain from learners in a corporate training environment who thought their e-learning's answer-until-correct questions were particularly worthless as a measure of their learning. They were given questions at the end of e-learning modules and could not complete the required module until they answered each question correctly. They received no penalty for guessing and no benefit for thoughtfully answering the question. It was pretty clear from interviewing learners that the answer-until-correct design was not adding to their learning. Perhaps this was more of a question of implementation rather than the answer-until-correct methodology itself. It is likely that a different implementation of the answer-until-correct design might be better received and more effective. For example, maybe learning would have benefitted more (a) if learners were given a higher score for first-try correct answers than for second-try answers, or (b) if the questions were more relevant to

the learners' work competence, or (c) if the e-learning content in general was more relevant to their work competence.

The Distinction between Building Understanding and Supporting Retrieval/Fluency

My distinction between building understanding and supporting retrieval/fluency has not been directly tested—as far as I can tell—in the academic literature. On the surface, the notion that learners are engaged in fundamentally different cognitive processes while building understanding compared to when they are practicing retrieval of well-learned information seems obvious.

When learners are well practiced in retrieving information, they can immediately process environmental cues and evoke a response. They may even limit attention toward other stimuli to enable quick responding. In comparison, when people are learning something new, their attention is more likely to wander to various cues so that they can make sense of what they are learning.

In relation to feedback, there are various hints in the research that the understanding-retrieval distinction has relevance. For example, Anderson and his colleagues (Anderson, Corbett, Koedinger, & Pelletier, 1995) have found that immediate feedback works well in their intelligent tutoring systems. This makes sense within the understanding-retrieval framework because the tutoring systems are helping the learners build understanding.

In the motor skills literature reviewed by Magill (1993), it appears that beginners—those who are by definition building understanding—benefit from significant amounts of feedback early in the process of learning. Biofeedback can be used to help people learn to monitor their heartbeats, a skill for which most people are beginners. Redundant feedback, although harmful for most learners, is beneficial for beginners because it enables them to get a feel for how their new movements are creating their effects. Wulf, Shea, and Matschiner (1998) found that the most complex skills benefited from more feedback, at least initially in the learning process.

While these results are consistent with my understanding-retrieval/fluency distinction, more research is needed.

***Complexity of Feedback Context Makes Research Difficult:
A Message for Researchers***

I started this report arguing that feedback has to be understood in all its complexity before practical recommendations can be fully appreciated. For researchers, who need to tease apart causal factors from extraneous factors, the task is even harder.

Considering All Phases of the Feedback Context

Researchers who study feedback need to take into consideration all phases of the feedback context. Below, I have outlined these phases as they move chronologically from left to right. I have also illustrated (with arrows moving right to left) how the feedback context can be considered to loop back to further learning opportunities.

Critical Variables to be Considered

As comprehensive reviews of the feedback literature make clear (see, for example, Mory, 2004), there are dozens of factors that impact feedback’s effects. Although I cannot hope to catalog all of these variables here, the following seem to be some of the most important.

Variable	Why Important
How experienced learners are with the concepts being learned. (Prelude to Phase L, as illustrated in the diagram above.)	Novices differ from experts in their ability to process new information in a given topic area.
How much the learners learn (or know) before they are given their initial retrieval practice (R_1). (See Phase L in the diagram above.)	The better learners learn, the more likely they’ll get answers correct, making feedback less important.
The activities/thinking (at I_1) that takes place in between the initial learning event (L) and the initial retrieval event (R_1).	Learners may reinforce what they learned or create interference in relation to what they learned.
The length of time (in I_1) between the initial learning event (L) and the initial retrieval event (R_1).	The longer the time, the more likely learners are to forget in the interval, causing the initial retrieval event to be less successful, causing more of a need for feedback.
Whether the learners are receiving their pre-feedback retrieval event (R_1) as they are “learning to understand” or “supporting future retrieval.”	If learners are learning to understand, retrieval acts to help build correct knowledge structures. If learners are practicing retrieval after they already understand the concepts, they are doing something fundamentally different ¹⁰ . Different feedback is likely to be required for each—for example, more extensive feedback to build understanding.

¹⁰ A reasonable analogy is learning the route to a new location. Initially, as you learn the route, you look around, consciously trying to build a correct mental model of the terrain, focusing on street signs, landmarks, and the relation between these many and various objects. Later, as you practice the route, you do very little of this mental map-building. Instead, you look for the landmarks you already know, and make quick decisions accordingly.

<p>The activities/thinking (at I_2) that takes place in between the initial retrieval event (R_1) and the feedback episode (F).</p>	<p>Between retrieval and feedback, learners may reinforce what they learned, encounter/seek information relevant to what they learned (info that acts essentially like feedback), or focus on information that will interfere with the targeted information. Non-intended feedback can negate or strengthen the experimental effects of intentional didactic feedback¹¹.</p>
<p>The length of time (during I_2) between the initial retrieval event (R_1) and the feedback episode (F).</p>	<p>This delay may affect the type of feedback that is needed or its relative effectiveness. For example, if learners are building understanding, a lengthy delay may be harmful as they struggle to remember the correct mental models they had previously built. However, a lengthy delay may benefit those supporting retrieval because of the benefits of spacing.</p>
<p>The type of feedback given. For example, how corrective it is, how empathetic, how long, how personalized, etc.</p>	<p>The type of feedback has obvious implications.</p>
<p>The activities/thinking (at I_3) that takes place in between the feedback episode (F) and the targeted retrieval event (R_2).</p>	<p>Between feedback and the final retrieval event, learners may reinforce what they learned, encounter/seek information relevant to what they learned (info that acts essentially like feedback), or focus on information that will interfere with the targeted information.</p>
<p>The amount of time during the final retention interval (at I_3) that takes place in between the feedback episode (F) and the targeted retrieval event (R_2).</p>	<p>Real-world learning usually requires a substantial “retention interval.” Research has clearly shown that a “spacing effect” helps enable learners to retain retrievability over time. Therefore, when we look at the timing of feedback—whether it is provided immediately or after some delay—it is critical to take this particular retention interval into account. Certainly, short retention intervals should not be used as the only measurement of feedback’s impact.</p>

¹¹ Non-intended feedback events can act as feedback. For example, after answering a question, a learner could try out his/her new knowledge on the job, thereby getting feedback.

Priorities for Further Research

Although research on feedback is in great abundance, as new discoveries of human learning have been uncovered, much of the research has been revealed as flawed. Particularly important due to current inadequacies, future research should utilize methodologies that test long-term retention. Final retention tests should be given at least a day after feedback (or additional retrieval practice), but preferably, retention tests should be delayed a week, month, or even longer to simulate real-world conditions.

These long-term retention tests should not be contaminated by immediate tests given to all comparison groups, because such immediate tests may equalize groups and thus diminish feedback effects. Of course, it is perfectly reasonable to study whether immediate tests have an effect by having some groups get immediate tests and having some groups not get immediate tests, but giving everyone an immediate test may hide the long-term feedback effects.

Studies of feedback should test long-term retention and not some measure of learning acquisition. As mentioned earlier, studies like those reviewed in Kulik and Kulik (1988) that used “the number of trials until acquisition” are particularly dangerous as they conflate repetition and feedback effects.

The notion that retrieval practice is essential following feedback should be tested further beyond the studies of Phye and Andre (1989) and Sturges (1972).

The contingencies regarding immediate and delayed feedback need further study. While the weight of research suggests that delayed feedback is better in most circumstances, the power of immediate feedback has not been definitively ruled out.

Similarly, further research is needed to confirm the proposal that a feedback spaced repetition effect is a main cause of the superiority of delayed over immediate feedback.

Further research is needed to determine the best feedback delays. Should we be delaying feedback for one hour, 24 hours, or what? Current research isn't clear in this regard.

Additional research is necessary to help us figure out whether learner control can help. There are a few studies on this, but it is too soon to draw any firm conclusions.

It appears that feedback is most effective when questions are repeated verbatim on a later test as compared with questions that are paraphrased or involve a deeper understanding of the material. This relationship is rather tenuous so it would be nice to figure out what, if anything, is going on.

Some authors have argued that feedback is ineffective when the material is very difficult and unnecessary when the material is very easy. Such connections need further study.

The answer-until-correct feedback methodology requires more study. Especially critical will be real-world research that varies the incentives to learners for taking each answer choice seriously.

The benefits that might accrue from tracking learner confidence require additional research.

Further research is needed to determine whether different types of learners require different types of feedback. Some authors have suggested some individual differences, but the research on this is extremely weak and tenuous.

It would be valuable to determine whether learners in realistic long-term learning experiences benefit from feedback that is varied in some way to maintain learner interest.

Feedback's effects on different types of learning materials should be studied, including differences between (a) declarative and procedural knowledge, (b) building understanding and supporting retrieval/fluency, (c) complex and simple, (d) learning situations and performance situations, etc.



Summarizing Principles and Recommendations (copied from Part 1)

1. The most important thing to remember about feedback is that it is generally beneficial for learners.
2. The second most important thing to remember about feedback is that it should be corrective. Typically, this means that feedback ought to specify what the correct answer is. When learners are still building understanding, however, this could also mean that learners might benefit from additional statements describing the “whys” and “wherefores.”
3. The third most important thing to remember about feedback is that it must be paid attention to in a manner that is conducive to learning.
4. Feedback works by correcting errors, whether those errors are detected or hidden.
5. Feedback works through two separate mechanisms: (a) supporting learners in correctly understanding concepts, and (b) supporting learners in retrieval.
6. Early in learning a topic, learners need to focus on building understanding. Later in learning a topic, learners need to focus on supporting long-term retrieval and fluency.
7. Complex concepts will usually require more time for learners to build understanding.
8. To help learners build understanding, feedback should diagnose learners’ incorrect mental models and specifically correct those misconceptions, thereby enabling additional correct retrieval practice opportunities.
9. To prepare learners for future long-term retrieval and fluency, learners need practice in retrieving. For this purpose, retrieval practice is generally more important than feedback.
10. Elaborative feedback may be more beneficial as learners build understanding, whereas brief feedback may be more beneficial as learners practice retrieval.
11. Immediate feedback prevents subsequent confusion and limits the likelihood for continued inappropriate retrieval practice.
12. Delayed feedback creates a beneficial spacing effect.
13. When in doubt about the timing of feedback, you can (a) give immediate feedback and then a subsequent delayed retrieval opportunity, (b) delay feedback slightly, and/or (c) just be sure to give some kind of feedback.

14. Feedback should usually be provided before learners get another chance to retrieve incorrectly again.
15. Provide feedback on correct responses when:
 - a. Learners experience difficulty in responding to questions or decisions.
 - b. Learners respond correctly with less-than-high confidence.
 - c. All the information learned is of critical importance.
 - d. Learners are relatively new to the subject material.
 - e. The concepts are very complex.
16. Provide feedback on incorrect responses:
 - a. Almost always.
 - b. Except:
 - i. When feedback would disrupt the learning event.
 - ii. When it would be better to wait to provide feedback.
17. When learners seek out and/or encounter relevant learning material either before or after feedback, this can modify the benefits of the feedback itself.
18. When learners are working to support retrieval or fluency, short-circuiting their retrieval practice attempts by enabling them to access feedback in advance of retrieval can seriously hurt their learning results.
19. When learners retrieve incorrectly and get subsequent well-designed feedback, they still have not retrieved successfully; so they need at least one additional opportunity to retrieve—preferably after a delay.
20. On-the-job support from managers, mentors, coaches, learning administrators, or performance-support tools can be considered a potentially powerful form of feedback.
21. Training follow-through software—that keeps track of learners’ implementation goals—provides another opportunity for feedback.
22. Feedback can affect future learning by focusing learners on certain aspects of learning material at the expense of other aspects of learning material. Learners may take the hint from the feedback to guide their attention in subsequent learning efforts.
23. Extra acknowledgements (when learners are correct) and extra handholding (when learners are wrong) are generally not effective (depending on the learners). In fact, when feedback encourages learners to think about how well they appear to be doing, future learning can suffer as learners aim to look good instead of working to build rich mental models of the learning concepts.

Reference List

- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463-470.
- Anderson, D. I., Magill, R. A., & Seklya, H. (2001) Motor learning as a function of KR schedule and the characteristics of task-intrinsic feedback. *Journal of Motor Behavior*, 33, 59-66.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.
- Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology*, 62, 148-156.
- Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*, 63, 186-188.
- Armstrong, T. R. (1970). *Feedback and perceptual-motor skill learning: A review of information feedback and manual guidance training techniques*. Technical Report No. 25, Human Performance Center, University of Michigan.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13, 111-127.
- Bahrlick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, 108, 296-308.
- Bahrlick, H. P., & Hall, L. K. (1991). Preventive and corrective maintenance of access to knowledge. *Applied Cognitive Psychology*, 5, 1-18.
- Bahrlick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 344-349.
- Bahrlick, H. P., & Phelps, E. (1988). The maintenance of marginal knowledge. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 178-192). Cambridge, UK: Cambridge University Press.
- Bahrlick, H. P., Bahrlick, L. E., Bahrlick, A. S., & Bahrlick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4, 316-321.

- 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*, 213-238.
- Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5*, 438-447.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues, Vol. 1., Memory in Everyday Life* (pp. 396-401). NY: Wiley.
- Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005). Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *Psychological Record, 55*, 401-418.
- Brosvic, G. M., Epstein, M. L., Dihoff, R. E., & Cook, M. J. (2006). Acquisition and retention of Esperanto: The case for error correction and immediate feedback. *Psychological Record, 56*, 205-218.
- Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514-527.
- Butler, A. C., & Roediger III, H. L. (in press for 2008). Feedback Enhances the Positive Effects and Reduces the Negative Effects of Multiple-Choice Testing. *Memory & Cognition*.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (in press for 2008a). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, in press.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (submitted for 2008b). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses.
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491-1494.
- Cain, L. F., & Willey, R. (1939). The effect of spaced learning on the curve of retention. *Journal of Experimental Psychology, 25*, 209-214.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). *Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis*. *Psychological Bulletin, 132*, 354-380.
- Clariana, R. B., Ross, S. M., & Morrison, G. R. (1991). The effects of different feedback strategies using computer-administered multiple-choice questions as instruction. *Educational Technology Research and Development, 39*, 5-17.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, 21, 451-467.

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

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 1, 309-330.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.) *Memory* (pp. 317-344). San Diego, CA: Academic Press.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2005). Adjunctive Role for Immediate Feedback in the Acquisition and Retention of Mathematical Fact Series by Elementary School Students Classified with Mild Mental Retardation. *Psychological Record*, 55, 39-66.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84, 795-805.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY, US, Random House.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5-12.

English, R. A., & Kinzer, J. R. (1966). The effect of immediate and delayed feedback on retention of subject matter. *Psychology in the Schools*, 3, 143-147.

Frase, L. T. (1967). Learning from prose material: Length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 58, 266-272.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 26, 1-104.

- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing and repetitions on recall and recognition. *Memory & Cognition*, 7, 95-112.
- Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392-399.
- Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior*, 10, 23-28.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Potomac, MD: Erlbaum.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562-567.
- Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.) *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (Volume 2, pp. 1-33). Hillsdale, NJ: Erlbaum.
- Jones, H. E. (1923-1924). Experimental studies of college teaching: The effect of examination on permanence of learning. *Archives of Psychology*, 10, 1-70.
- Kang, S.H.K., McDermott, K.B. & Roediger, H.L., III (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *The European Journal of Cognitive Psychology*, 19, 528-558.
- Karraker, R. J. (1967). Knowledge of results and incorrect recall of plausible multiple-choice alternatives. *Journal of Educational Psychology*, 58, 11-14.
- Keenan, V. C., & Langer, P. (1993). Instructional Feedback in Context. *Psychological Reports*, 73, 107-112.
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232.
- Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, 63, 505-512.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279-308.

- Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522-528.
- Kulik, J. A., & Kulik, C-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79-97.
- Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109*, 451-464.
- Lavery, J. J. (1962). Retention of simple motor skills as a function of type of knowledge of results. *Canadian Journal of Psychology, 16*, 300-311.
- Lavery, J. J. (1964). Retention of a skill following training with and without instructions to retain. *Perceptual and Motor Skills, 18*, 275-281.
- Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Different effects for discrete and continuous tasks. *Research Quarterly for Exercise and Sport, 60*, 59-65.
- Magill, R. (1993). Augmented feedback in skill acquisition. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of Research on Sport Psychology* (pp. 193-212). NY: Macmillan.
- 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.
- Modigliani, V., & Hedges, D. G. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 426-436.
- Moore, J. W., & Smith, W. I. (1961). Knowledge of results in self-teaching spelling. *Psychological Reports, 9*, 717-726.
- More, A. (1969). Delay of feedback and the acquisition and retention of verbal materials in the classroom. *Journal of Educational Psychology, 60*, 339-342.
- Mory, E. H. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Technology Research and Development, 40*, 5-20.
- Mory, E. H. (2004). Feedback Research Revisited. In Jonassen, David H. (Ed), *Handbook of Research on Educational Communications and Technology (2nd ed.)*. (pp. 745-783). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Newman, M. I., Williams, R. G., & Hiller, J. H. (1974). Delay of information feedback in an applied setting: Effects on initially learned and unlearned items. *Journal of Experimental Education, 42*, 55-59.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*, 18-22.

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*, 3-8.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*, 187-193.

Peeck, J., van den Bosch, A. B., & Kreupeling, W. J. (1985). Effects of informative feedback in relation to retention of initial responses. *Contemporary Educational Psychology, 10*, 303-313.

Phe, G. D. (1991). Advice and feedback during cognitive training: Effects at acquisition and delayed transfer. *Contemporary Educational Psychology, 16*, 87-94.

Phe, G. D., & Andre, T. (1989). Delayed retention effect: Attention, perseveration, or both. *Contemporary Educational Psychology, 14*, 173-185.

Phe, G. D., & Sanders, C. E. (1994). Advice and feedback: Elements of practice for problem solving. *Contemporary Educational Psychology, 19*, 286-301.

Pressey, S. L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *The Journal of Psychology, 29*, 417-447.

Rea, C. P., & Modigliani, V. (1988). Educational implications of the spacing effect. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.) *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 402-406). New York: John Wiley & Sons.

Renner, K. E. (1964). Delay of reinforcement: A historical review. *Psychological Bulletin, 61*, 341-361.

Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research, 49*, 181-196.

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

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

- Ross, R. T., & Pirie, M. (1934). The persistence of errors in successive true-false tests. *Journal of Educational Psychology, 25*, 422-426.
- Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3*, 241-249.
- Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written material. *Journal of Educational Psychology, 58*, 56-61.
- Rothstein, A. L., & Arnold, R. K. (1976). Bridging the gap: Application of research on video-tape feedback and bowling. *Motor Skills: Theory into Practice, 1*, 35-62.
- Ruch, T. C. (1928). Factors influencing the relative economy of massed and distributed practice in learning. *Psychological Review, 35*, 19-45.
- Sagerman, N., & Mayer, R. E. (1987). Forward transfer of different reading strategies evoked by adjunct questions in science text. *Journal of Educational Psychology, 79*, 189-191.
- Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*, 355-386.
- Sassenrath, J. M. (1975). Theory and results on feedback and retention. *Journal of Educational Psychology, 67*, 894-899.
- Sassenrath, J. M., & Gaverick, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology, 56*, 259-263.
- Sassenrath, J. M., & Yonge, G. D. (1968). Delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology, 59*, 69-73.
- Sassenrath, J. M., & Yonge, G. D. (1969). Effects of delayed information feedback and feedback cues in learning on delayed retention. *Journal of Educational Psychology, 60*, 174-177.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
- Shrock, S. A., & Coscarelli, W. C. (2007). *Criterion-referenced Test Development: Technical and Legal Guidelines for Corporate Training, 3rd Edition*. San Francisco: Pfeiffer.
- Skinner, B. F. (1954). *The science of learning and the art of teaching*. Harvard Educational Review, 24, 86-97.

- Skinner, B.F. (1968). *The technology of teaching*. East Norwalk, CT, US, Appleton-Century-Crofts.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641-656.
- Sturges, P. T. (1969). Verbal retention as a function of the informativeness and delay of informative feedback. *Journal of Educational Psychology*, *60*, 11-14.
- Sturges, P. T. (1972). Information delay and retention: Effect of information in feedback and tests. *Journal of Educational Psychology*, *63*, 32-43.
- Sturges, P. T. (1978). Delay of informative feedback in computer-assisted testing. *Journal of Educational Psychology*, *70*, 378-387.
- Surber, J. R., & Anderson, R. C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology*, *67*, 170-173.
- Swinnen, S. P. (1990). Interpolated activities during the knowledge-of-results delay and post-knowledge-of-results interval: Effects on performance and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 692-705.
- Swinnen, S. P., Schmidt, R. A., Nicholson, D. E., & Shapiro, D. C. (1990). Information feedback for skill acquisition: Instantaneous knowledge of results degrades learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 706-716.
- Thalheimer, W. (2006, February). *Spacing Learning Events Over Time: What the Research Says*. Retrieved January 9, 2008, from <http://www.work-learning.com/catalog/>.
- Thalheimer, W. (2007, April). *Measuring Learning Results: Creating fair and valid assessments by considering findings from fundamental learning research*. Retrieved January 9, 2008, from <http://www.work-learning.com/catalog/>.
- Thomas, R. E., Croal, B. L., Ramsay, C., Eccles, M., & Grimshaw, J. (2006). Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet*, *367*, 1990-1996.
- Toppino, T. C., & Brochin, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research*, *83*, 119-124.
- Webb, J. M., Stock, W. A., & McCarthy, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology*, *19*, 251-265.
- Wulf, G., Shea, C. H., & Matschiner, S. (1998). Frequent feedback enhances complex motor skill learning. *Journal of Motor Behavior*, *30*, 180-192.