

The hypercorrection effect persists over a week, but high-confidence errors return

Andrew C. Butler · Lisa K. Fazio · Elizabeth J. Marsh

Published online: 12 October 2011
© Psychonomic Society, Inc. 2011

Abstract People's knowledge about the world often contains misconceptions that are well-learned and firmly believed. Although such misconceptions seem hard to correct, recent research has demonstrated that errors made with higher confidence are more likely to be corrected with feedback, a finding called the *hypercorrection effect*. We investigated whether this effect persists over a 1-week delay. Subjects answered general-knowledge questions about science, rated their confidence in each response, and received correct answer feedback. Half of the subjects reanswered the same questions immediately, while the other half reanswered them after a 1-week delay. The hypercorrection effect occurred on both the immediate and delayed final tests, but error correction decreased on the delayed test. When subjects failed to correct an error on the delayed test, they sometimes reproduced the same error from the initial test. Interestingly, high-confidence errors were more likely than low-confidence errors to be reproduced on the delayed test. These findings help to contextualize the hypercorrection effect within the broader memory literature by showing that high-confidence errors are more likely to be corrected, but they are also more likely to be reproduced if the correct answer is forgotten.

Keywords Cued recall · Knowledge · Memory · Metamemory

A. C. Butler (✉) · E. J. Marsh
Psychology & Neuroscience, Duke University,
P.O. Box 90086, Durham, NC 27708-0086, USA
e-mail: andrew.butler@duke.edu

L. K. Fazio
Department of Psychology, Carnegie Mellon University,
5000 Forbes Ave,
Pittsburgh, PA 15213, USA

People acquire misconceptions about the world from many different sources, including fictional stories (Marsh & Fazio, 2007), popular history films (Butler, Zaromb, Lyle, & Roediger, 2009), and multiple-choice tests (Roediger & Marsh, 2005). Often this false knowledge is innocuous, but sometimes it undermines our understanding of the world (Hammer, 1996). A prime example comes from the educational video *A Private Universe* (Schneps, 1989), in which recent graduates of Harvard University were asked to explain what causes seasonal changes in the Earth's climate. Almost all of the graduates incorrectly attributed seasonal changes to fluctuations in the distance between the Earth and the Sun (the seasons are actually caused by the tilt of the Earth's rotational axis). Interestingly, the graduates retrieved their answers quickly and reported them with confidence, which suggests that this particular misconception was firmly entrenched in their knowledge. Given the prevalence of such misconceptions, it is critical to determine how best to correct errors in knowledge, especially when these errors are well-learned and produced with high confidence.

One method of correcting errors in knowledge is to provide feedback. Although feedback is often very effective (see Hattie & Timperley, 2007; Shute, 2008), prior research has suggested that it should be extremely difficult to correct misconceptions that are highly accessible in memory and strongly believed. In the substantial literature on proactive interference, a common finding is that well-learned information interferes with the acquisition of related information (see Anderson & Neely, 1996; Postman, 1976). Everyday experience also suggests that it is particularly hard to correct a well-learned error, such as when one must learn to correctly pronounce a person's name after mispronouncing the name for a long period of time.

Contrary to these predictions, recent studies have found that errors made with high confidence are more likely to be

corrected with feedback than are low-confidence errors, a finding called the *hypercorrection effect* (Butterfield & Metcalfe, 2001). The hypercorrection effect is a highly replicable phenomenon that has been independently observed in experiments conducted by several different research groups (Butler, Karpicke, & Roediger, 2008; Butterfield & Metcalfe, 2006; Fazio & Marsh, 2009, 2010; Kulhavy, Yekovich, & Dyer, 1976). For example, Butterfield and Metcalfe (2001) gave subjects a short-answer test that consisted of various general-knowledge questions. After responding to each question, the subjects rated their confidence on a 7-point scale and then received feedback (either “you’re right” or the correct answer, if they were wrong). After a 5-min delay, they were retested on the questions. Errors given a higher confidence rating on the initial test were more likely to be corrected on the final test than were errors previously made with lower confidence.

How can this apparent contradiction between the hypercorrection effect and existing theories of memory be resolved? The key to understanding how the hypercorrection effect fits within the broader memory literature may be the time scale over which the phenomenon occurs. Almost all of the studies on the hypercorrection effect have used a short delay between the presentation of feedback and the final test (e.g., 5 min). Testing memory after such a brief retention interval may not accurately assess whether these errors will remain corrected over longer periods of time. Interference theory would predict that the hypercorrection effect is a relatively short-lived phenomenon; high-confidence errors may be more likely to be corrected initially, but gradually these prepotent errors will return and interfere with memory for the correction. For example, proactive interference tends to be minimal when memory is tested immediately, but increases steadily as a function of delay (e.g., Briggs, 1954).

Yet there is one piece of evidence to suggest that the hypercorrection effect persists over longer periods of time: Butterfield and Mangels (2003, Exp. 2) found that high-confidence errors were more likely to be corrected on the final test, regardless of whether it was given immediately or after a 1-week delay. Nevertheless, it is difficult to draw firm conclusions from this study because of two issues. First, subjects produced a relatively low number of high-confidence errors on the initial test, which increases the likelihood that the finding was due to random variation. Second, this study did not assess whether subjects produced the same error on the final test or a different error—all errors were treated the same. Thus, it is unclear whether high-confidence errors were more likely to be reproduced on the final test relative to errors made with lower confidence (e.g., a low-confidence error might have switched to a different incorrect response).

The main goal of the present research was to gain a better understanding of how the relationship between

response confidence and error correction changes over time. Of specific interest was whether the correction of high-confidence errors would persist over a longer period of time or whether these errors would gradually return. To answer this question, we conducted an experiment in which subjects answered general-knowledge questions and rated their confidence in each answer, receiving feedback on their responses. Critically, one group of subjects reanswered the questions after a delay of 6 min, whereas a second group reanswered the questions after 1 week. We used general-knowledge questions about scientific facts from biology, physics, astronomy, and other fields. In order to increase the number of high-confidence errors made by subjects, we included many questions that probed common misconceptions about science, such as the aforementioned example about the cause of seasonal changes in Earth’s climate. As a secondary goal, we wanted to explore whether subjects would remember their errors and whether memory for these errors helped or hindered subsequent error correction. To this end, we asked subjects to recall their response on the initial test after they had finished answering all of the questions on the final test.

Method

Subjects and design

A group of 50 undergraduates at Duke University participated for course credit or pay. The experiment had one independent variable (retention interval: 6 min, 1 week), which was manipulated between subjects.

Materials

The materials consisted of 120 general-knowledge questions about science (e.g., *What is stored in a camel’s hump?* Answer: *Fat*; *How many chromosomes do humans have?* Answer: *46*; *What is the driest area on Earth?* Answer: *Antarctica*). Each response consisted of a single word or a short phrase. The questions were generated from a variety of Internet websites, including Wikipedia (wikipedia.org), Discovery Channel (www.discovery.com), and Science Hobbyist (www.amasci.com). Every fact was verified by consulting additional sources. Piloting was conducted to ensure that the questions would yield sufficient variability in terms of both accuracy and confidence ratings (i.e., a good distribution of responses at all levels of the confidence scale).

Procedure

Upon arriving in the laboratory for the experiment, subjects were randomly assigned to one of the two retention interval

groups. The entire experiment was conducted on the computer. First, subjects took a short-answer test that consisted of 120 questions. They were instructed to provide a response for every question, even if they had to guess (i.e., forced report). After entering their response to each question, they rated their confidence on a scale of 1 (*sure wrong*) to 7 (*sure correct*), and then they received feedback. Feedback consisted of the presentation of the correct answer for 6 s, and it was given for both correct and incorrect responses. Subjects were told to study the feedback because they would be tested again on the material later. After performing a filler task (visuospatial puzzles), subjects reanswered the same 120 questions either immediately (6-min retention interval group) or 1 week later (1-week retention interval group). Like the initial test, the final test was in short-answer format and was self-paced and forced report. However, unlike the initial test, subjects did not receive feedback. After taking the final test, subjects were re-presented with each question and asked to recall their response on the initial test. If they could not recall their initial response, they were instructed to guess.

Results

All results, unless otherwise stated, were significant at the .05 level. Eta-squared and Cohen's d are the measures of effect size reported for all significant effects in the ANOVA and the t -test analyses, respectively.

Coding

Two coders independently scored all of the responses. Both coders were blind to condition, and they scored all the responses for a given question together to increase consistency. Cohen's kappa was calculated to assess interrater reliability. Reliability was very high ($\kappa = .97$), and the first author (A.C.B.) resolved the few disagreements.

Initial test

The proportion of correct responses on the initial test was relatively low (grand mean = .38), which was desirable for investigating error correction. As expected, there was no significant difference in performance between the two retention interval groups because the manipulation had yet to be implemented ($t < 1$). In order to investigate the relationship between response confidence and error correction, it is important to have a good distribution of responses across the confidence scale. Table 1 shows the numbers of responses on the initial test as a function of confidence rating, response outcome (correct/incorrect), and retention interval group. Clearly, there were large numbers of correct and incorrect responses at every level of confidence, and many high-confidence errors for both groups.

Final test

As expected, subjects used the feedback to correct errors made on the initial test, which led to an increase in the proportions of correct responses on the final test. The subjects in the 6-min retention interval group produced a significantly greater proportion of correct responses than did the subjects in the 1-week retention interval group [.90 vs. .71; $t(48) = 6.11$, standard error of the difference (SED) = .03, $d = 1.31$].

Conditional analyses

Several conditional analyses were conducted on the data. The first set of analyses investigated the relationship between response confidence and error correction. Figure 1 depicts the mean proportions of errors on the initial test that were corrected on the final test as a function of initial test confidence and retention interval group. As the figure shows, the greater the confidence in the error, the more likely it was to be corrected—a hypercorrection effect. Importantly, this relationship held for both retention intervals, replicating Butterfield and Mangels (2003) and

Table 1 The numbers of responses on the initial test as a function of confidence rating, response outcome (correct/incorrect), and retention interval group

Response Outcome	Retention Interval Group	Confidence Rating							Total
		1	2	3	4	5	6	7	
Correct	6 Min	46	30	39	139	119	155	554	1,082
	1 Week	50	41	52	130	157	187	564	1,181
	Total	96	71	91	269	276	342	1,118	2,263
Incorrect	6 Min	614	159	158	440	222	118	207	1,918
	1 Week	536	229	188	346	200	119	201	1,819
	Total	1,150	388	346	786	422	237	408	3,737
Grand total		1,246	459	437	1,055	698	579	1,526	6,000

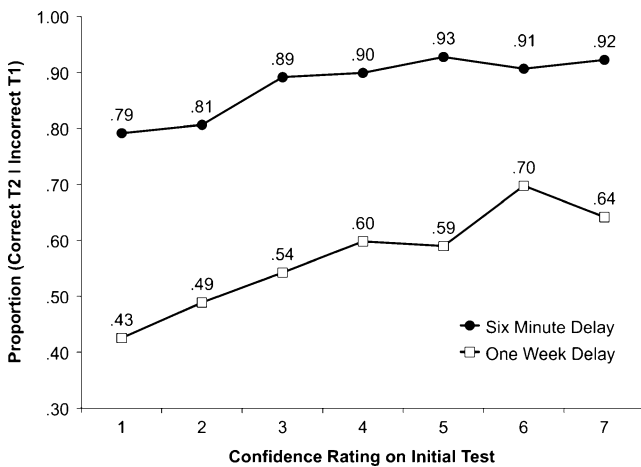


Fig. 1 Mean proportions of errors on the initial test that were corrected on the final test, as a function of confidence on the initial test and retention interval group

providing additional evidence that the hypercorrection effect persists over longer time periods. For each subject, we computed a within-subjects Goodman–Kruskal gamma correlation (Goodman & Kruskal, 1954) between the confidence ratings on the initial test and the response outcomes on the final test (correct or incorrect) for errors on the initial test (gamma is a nonparametric statistic commonly used in the metacognition literature to deal with ordinal-scale data). A one-sample *t*-test confirmed that the mean gamma correlations for the 6-min [$M = .28$; $t(24) = 4.22$, $SEM = .07$, $d = 0.86$] and 1-week [$M = .19$; $t(24) = 3.95$, $SEM = .05$, $d = 0.79$] retention groups were each significantly different from zero. However, an independent-samples *t*-test comparing the two groups showed that the two mean gamma correlations did not significantly differ from each other [.28 vs. .19; $t(47) = 1.13$, $SED = .08$, $p = .27$].

Although the hypercorrection effect remained when memory was tested after a longer retention interval, there was a significant decrease in the overall proportions of errors corrected [.86 vs. .56; $t(48) = 9.13$, $SED = .03$, $d = 1.58$], indicating that subjects forgot many of the correct responses over the 1-week delay. This result is important because it raises the question of whether the original errors returned. Figure 2 depicts the mean proportions of errors on the initial test for which the same error was produced on the final test as a function of confidence on the initial test and of retention interval group. In the 6-min retention interval group, subjects rarely produced the same error on the final test, regardless of their initial confidence in the error. However, subjects in the other group began to reproduce their errors from the initial test after a 1-week delay. Importantly, the greater their original confidence in the error, the more likely they were to reproduce it on the final test 1 week later. We computed a gamma correlation for each subject between confidence rating on the initial test

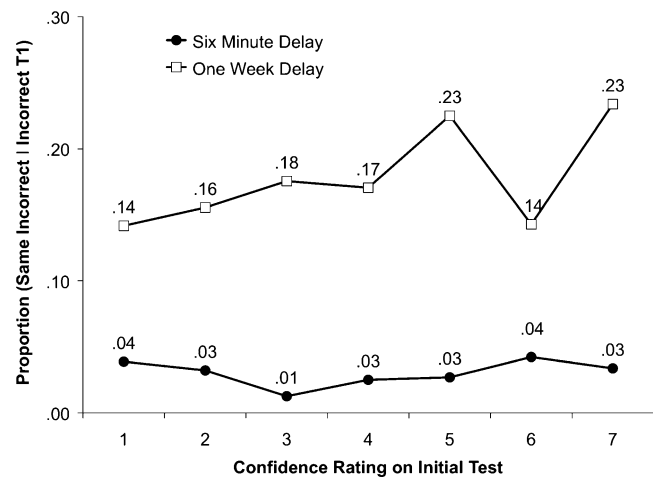


Fig. 2 Mean proportions of errors on the initial test for which the same error was produced on the final test, as a function of confidence on the initial test and retention interval group

and response outcome on the final test (same error or other outcome) for items that were incorrect on the initial test. A one-sample *t*-test confirmed that the mean gamma correlation was significantly different from zero [$M = .15$; $t(24) = 2.53$, $SEM = .06$, $d = 0.51$]. We did not conduct the equivalent analysis for the 6-min retention interval condition because not enough observations were produced (i.e., same-error responses) to compute the gamma correlations.

Recall of initial test responses

A final set of analyses explored subjects' accuracy in recalling their initial responses and whether memory for an initial error facilitated or interfered with subsequent error correction. When subjects had answered the initial questions correctly, they almost always correctly remembered their initial response when asked after the final test (grand mean = .98). They were also very good at recalling their errors on the initial test. However, subjects in the 6-min retention interval group were significantly more accurate than subjects in the 1-week retention interval group [.85 vs. .61; $t(48) = 5.54$, $SED = .04$, $d = 1.24$]. Gamma correlations were computed for each subject in order to assess the relationship between the accuracy of initial error recall and confidence rating. The higher the subjects' confidence in the error, the more likely they were to recall it later, and this relationship was present in both retention interval groups. A one-sample *t*-test confirmed that the mean gamma correlations were significantly different from zero in both the 6-min delay group [$M = .45$; $t(23) = 8.32$, $SEM = .05$, $d = 1.67$] and the 1-week delay group [$M = .47$; $t(24) = 12.07$, $SEM = .04$, $d = 2.47$].

Since subjects could remember many of their errors from the initial test, a follow-up analysis was conducted to investigate how memory for an error affected performance on the final test.

For each subject, we calculated the proportions of initial errors that were corrected on the final test as a function of whether or not the initial test error was recalled after the final test. Collapsing across retention interval groups, subjects were more likely to correct their errors if they remembered the error ($M = .72$) than if they did not remember it ($M = .65$). This observation was confirmed by a 2 (error recall) \times 2 (retention interval) repeated measures ANOVA, which revealed significant main effects of error recall [$F(1, 47) = 7.04$, $MSE = .11$, $\eta^2 = .13$] and retention interval [$F(1, 47) = 80.54$, $MSE = 2.57$, $\eta^2 = .63$]. However, the interaction was not significant ($F < 1$).

Discussion

The present study yielded several interesting findings that enhance our understanding of the relationship between response confidence and error correction. We found a hypercorrection effect on a final test given after 6 min, replicating many previous studies (Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009, 2010; Kulhavy et al., 1976). However, more importantly, the effect persisted over a 1-week delay, replicating Butterfield and Mangels (2003). Despite the persistence of the hypercorrection effect, we also found that subjects forgot many of the correct responses over the 1-week delay and began to reproduce their errors from the initial test. A conditional analysis revealed an interesting new finding: The greater the confidence in an error on the initial test, the more likely it was to be reproduced on the final test 1 week later. Finally, there was some evidence that subjects were more likely to correct their errors if they remembered the error.

Our study helps to situate the hypercorrection effect within the broader memory literature. The finding that high-confidence errors are more likely to be corrected with feedback seems to contradict current theories of memory, which predict substantial proactive interference when people try to correct misconceptions that are deeply entrenched in knowledge (see Anderson & Neely, 1996; Postman, 1976). However, the results of the present experiment suggest a potential resolution to this apparent contradiction: Although high-confidence errors are more likely to be corrected, they are also more likely to be reproduced if the correct answer is forgotten. Thus, a shift occurs gradually over time as correct answers are forgotten and high-confidence errors return. This characterization of the hypercorrection effect dovetails nicely with the proactive-interference literature. Proactive interference increases as a function of the retention interval—initially, the newly learned response (the correct answer) is more likely to be retrieved, but over time the

original response (the error) returns to being the dominant response (e.g., Briggs, 1954).

With respect to a theoretical explanation for our finding, the “new theory of disuse” proposed by Bjork and Bjork (1992) provides a useful framework. Reviving a once-influential idea that had been largely forgotten (e.g., Estes, 1955), Bjork and Bjork made the distinction between the *storage strength* and *retrieval strength* of representations in memory. *Storage strength* refers to how well a piece of information is learned, whereas *retrieval strength* reflects the momentary accessibility of that information. According to their theory, both storage strength and retrieval strength increase with each exposure to the information (e.g., a study trial or retrieval from memory). However, while the accumulated storage strength is never lost, retrieval strength decreases over time, due to interference from subsequent exposure to other pieces of information. We can interpret the present findings with this distinction in mind. High-confidence errors should have high storage and retrieval strength. The presentation of feedback after an error should reduce the retrieval strength of the error while increasing the retrieval strength of the correct answer. Thus, when memory is tested soon after feedback, the correct answer will be retrieved. However, the storage strength of the correct answer should be much lower than that of the high-confidence error because of the difference in the number of prior exposures to these two pieces of information. As retrieval strength is lost over time, the high-confidence error will be more likely to be retrieved because of this difference in the underlying storage strength.

The foregoing analysis provides the motivation for our interest in whether people’s memories for their errors play a role in error correction. As the results of the present experiment demonstrate and the “new theory of disuse” predicts (Bjork & Bjork, 1992), people are quite good at recalling errors that they made earlier, especially high-confidence errors (see, too, Peeck & Tillema, 1979). One might expect that if people can remember their initial error, it would interfere with their ability to remember the correct response (e.g., due to response competition). Indeed, the assumption that forgetting an error facilitates learning of the correct response is the dominant explanation for the finding that delayed feedback produces better retention than does immediate feedback (i.e., the interference-perseveration hypothesis; Kulhavy & Anderson, 1972). However, contrary to this assumption, we found a small but significant advantage in error correction when subjects could accurately remember the error that they had produced on the initial test. Perhaps remembering a prior error can facilitate error correction, which might occur by tagging the error response as incorrect (for a similar idea, see Schwarz, Sanna, Skurnik, & Yoon, 2007) or associating the correct response with the

error response to form a mediational chain (e.g., Russell & Storms, 1955). Clearly, this idea is somewhat speculative and necessitates further research, but we have found some preliminary evidence to support it.

On a final note, our findings have implications for educators who strive to correct misconceptions that are firmly entrenched in knowledge. Given the prevalence of misconceptions about science (e.g., Hammer, 1996) and their persistence despite instruction to the contrary (e.g., Gutman, 1979), it is critical to determine how to effectively teach people the correct information. Although our findings suggest that one presentation of feedback is not enough to produce a lasting correction of high-confidence errors, we want to stress that this does not undermine the practical importance of the hypercorrection effect. Rather, we think that the hypercorrection effect provides a valuable opportunity. Our findings emphasize the need to capitalize on the hypercorrection effect before high-confidence errors return by providing additional opportunities to learn the correct information.

One potential solution would be to refute the misconception (e.g., Kowalski & Taylor, 2009) and then to provide practice retrieving the correct response from memory in order to increase long-term retention (e.g., Butler, 2010; see Roediger & Butler, 2011). Recent research by Fazio (2011) has suggested that this procedure might be highly effective. In her experiment, subjects took an initial general-knowledge test, received feedback on all of their responses, and then immediately retook the same test. This immediate retest gave subjects the opportunity to practice retrieving the correct answers that had been provided in the feedback. On a final test given 1 week later, performance showed hardly any forgetting—subjects retained almost of all of the correct responses and rarely reproduced their initial test errors. In other words, retesting subjects soon after they have received feedback on a high-confidence error increased retention for the correct response, and thus produced a long-lasting correction.

Our findings and those of Fazio (2011) are complementary, in that together they demonstrate the importance of providing additional opportunities to learn the correct information. As our study shows, when no further practice is provided, high-confidence errors are more likely to return over time. However, when additional practice is provided, as in Fazio's study, high-confidence errors remain corrected. Although such additional practice could take many forms, we think that repeated retrieval practice with feedback after each attempt might be the most effective intervention (see Roediger & Butler, 2011). In sum, the process of correcting misconceptions in knowledge will not be quick or easy, but the hypercorrection effect could serve as the foundation from which to build long-term retention of correct information.

Author note This research was supported by a Collaborative Activity Award from the James S. McDonnell Foundation's 21st Century Science Initiative in Bridging Brain, Mind and Behavior (to E.J.M.). The authors thank Aaron Johnson and Gene Eng for their help creating the materials, programming, collecting the data, and coding. The authors also thank the Marsh Lab and the Memory Reading Group at the University of North Carolina for feedback on earlier drafts of the manuscript.

References

- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Vol. 10. Memory* (2nd ed., pp. 237–313). San Diego: Academic Press.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in Honor of William K. Estes, vol. 2* (pp. 35–67). Hillsdale, NJ: Erlbaum.
- Briggs, G. E. (1954). Acquisition, extinction, and recovery functions in retroactive inhibition. *Journal of Experimental Psychology*, *47*, 285–293. doi:10.1037/h0060251
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 918–928. doi:10.1037/0278-7393.34.4.918
- Butler, A. C., Zaromb, F., Lyle, K. B., & Roediger, H. L., III. (2009). Using popular films to enhance classroom learning: The good, the bad, and the interesting. *Psychological Science*, *20*, 1161–1168.
- Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, *17*, 793–817. doi:10.1016/S0926-6410(03)00203-9
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1491–1494.
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition Learning*, *1*, 69–84.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369–377. doi:10.1037/h0046888
- Fazio, L.K. (2011). *Persistence of the hypercorrection effect: Do high-confidence errors remain corrected over a delay?* (in press).
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*, 88–92. doi:10.3758/PBR.16.1.88
- Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science*, *21*, 801–803.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764.
- Gutman, A. (1979). Misconceptions of psychology and performance in the introductory course. *Teaching of Psychology*, *6*, 159–161.
- Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, *64*, 1316–1325.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology*, 36, 153–159.
- Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, 63, 505–512.
- Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology*, 68, 522–528.
- Marsh, E. J., & Fazio, L. K. (2007). Learning from fictional sources. In J. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 397–413). New York: Psychology Press.
- Peeck, J., & Tillema, H. H. (1979). Delay of feedback and retention of correct and incorrect responses. *The Journal of Experimental Education*, 47, 171–178.
- Postman, L. (1976). Interference theory revisited. In J. Brown (Ed.), *Recall and recognition* (pp. 157–181). New York: Wiley.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27. doi:10.1016/j.tics.2010.09.003
- Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1155–1159. doi:10.1037/0278-7393.31.5.1155
- Russell, W. A., & Storms, L. H. (1955). Implicit verbal chaining in paired-associate learning. *Journal of Experimental Psychology*, 49, 287–293.
- Schneps, M. (1989). *A private universe [Video]*. Santa Monica, CA: Pyramid.
- Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, vol. 39 (pp. 127–161). San Diego: Academic Press. doi:10.1016/S0065-2601(06)39003-X
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.