Mem Cogn
DOI 10.3758/s13421-014-0441-y

Waiting for feedback helps if you want to know the answer: the role of curiosity in the delay-of-feedback benefit

Kellie M. Mullaney · Shana K. Carpenter · Courtney Grotenhuis · Steven Burianek

Abstract When participants answer a test question and then receive feedback of the correct answer, studies have shown that the feedback is more effective when it is delayed by several seconds rather than provided immediately (e.g., Brackbill & Kappy, *Journal of Comparative and Physiological Psychology, 55*, 14–18, 1962; Schroth, *Contemporary Educational Psychology, 17*, 78–82, 1992). Despite several demonstrations of this *delay-of-feedback benefit*, a theoretical explanation for this finding has not yet been developed. The present study tested the hypothesis that brief delays of feedback are beneficial because they encourage anticipation of the upcoming feedback. In Experiment 1, participants answered obscure trivia questions, and before receiving the answer, they rated their curiosity to know the answer. The answer was then provided either immediately or after a 4-s delay. A later final test over the same questions revealed a significant delay-of-feedback benefit, but only for items that had been rated high in curiosity. Experiment 2 replicated this same effect and showed that the delay-of-feedback benefit only occurs when feedback is provided after a variable, unpredictable time duration (either 2, 4, or 8 s) rather than after a constant duration (always 4 s). These findings demonstrate that the delay-of-feedback effect appears to be greatest under conditions in which participants are curious to know the answer and when the answer is provided after an unpredictable time interval.

Keywords Delay of feedback · Curiosity · Memory

Feedback is essential to learning. Research has shown that when participants answer a question incorrectly on a memory test, they are very unlikely to correct this error unless they are provided with knowledge of the correct response (e.g., Carpenter, Sachs, Martin, Schmidt, & Looft, 2012; Fazio, Huelser, Johnson, & Marsh, 2010; Finn & Metcalfe, 2010; Kang et al., 2011; Lhyle & Kulhavy, 1987; Pashler, Cepeda, Wixted, & Rohrer, 2005). A related topic concerns the *timing* of feedback. When participants answer a question incorrectly and then receive a presentation of the correct response, when should this presentation occur? Is there an optimal time during which to provide feedback? Or, is the timing unimportant as long as feedback is eventually provided?

An intuitive assumption is that feedback should be provided as soon as possible. Indeed, decades of research on operant learning have demonstrated that behavior is most effectively reinforced when knowledge of the desired response is provided immediately after the response occurs (e.g., Perin, 1943; Pubols, 1958; Tarpy & Sawabini, 1974). Interestingly however, research using cognitive tasks has found that immediate feedback is *not* always best. Instead, performance can be enhanced by delaying feedback for brief periods of time (i.e., up to several seconds) after a response is made.

For example, Brackbill and colleagues found that children's performance in a visual discrimination task was enhanced under conditions in which the feedback presentation was delayed by several seconds rather than provided immediately (e.g., Brackbill, 1964; Brackbill, Bravos, & Starr, 1962; Brackbill, Isaacs, & Smelkinson, 1962). In one study (Brackbill & Kappy, 1962), children were presented with pairs of line drawings (e.g., a boat and a star) and had to guess which drawing was correct. After the child chose one of the drawings, feedback (in the form of a light flashing above the correct drawing) was provided either immediately or after a 10-s delay. Although immediate feedback led to faster acquisition—that is, fewer trials required to reach a criterion of three correct responses during learning—it actually led to *worse* performance on a later test of relearning. On this later test, children who had previously received 10-s delayed

K. M. Mullaney · S. K. Carpenter (✉) · C. Grotenhuis · S. Burianek
Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011, USA
e-mail: shacarp@iastate.edu

Published online: 03 July 2014

feedback reached a criterion of three correct responses in fewer trials than did children who had previously received immediate feedback.

Similarly, Schroth (1992) observed that initial acquisition in a concept-learning task was hindered when feedback was provided after a delay of 30 s rather than immediately (requiring 48 vs. 24 trials, respectively, to reach a criterion of 90 %). Later performance on the same task, however, revealed a significant benefit of 30-s delayed feedback over immediate feedback (requiring 30 vs. 45 trials, respectively, to reach criterion). In a similar study involving discovery of artificial grammar rules (Schroth & Lund, 1993), participants learned the task faster if they received immediate feedback rather than 30-s delayed feedback (requiring 19 vs. 40 trials, respectively, to reach the criterion of 90 %), but later retention of these rules was better for items that were given 30-s delayed feedback rather than immediate feedback (72 % vs. 47 % accuracy, respectively).

Other studies using different types of memory tasks have confirmed that delayed feedback can benefit retention more than does immediate feedback. Sturges, Sarafino, and Donaldson (1968) presented children with a U.S. state and two cities, and asked them to guess which city was the capital. Feedback of the correct answer was given either immediately or after a 10-s delay. Although immediate feedback led to faster acquisition than did delayed feedback (approximately 4.5 vs. 8 errors, respectively), delayed feedback led to better performance than did immediate feedback on a later test of relearning (approximately 2 vs. 4.5 errors, respectively). Although the studies reviewed so far involved multitrial learning to criterion, Sassenrath and Yonge (1969) also found that on a single attempt at answering a multiple-choice question, participants who received feedback of the correct answer after a 10-s delay retained the material better than those who received feedback immediately.

The beneficial effect of delayed over immediate feedback, hereafter referred to as the *delay-of-feedback benefit*, has been demonstrated fairly consistently. However, a theoretical explanation for this effect has not yet been developed. Why is it beneficial to delay feedback for brief periods of time? In the studies discussed so far, it is possible that this effect was driven by the fact that participants spent more time on delayed-feedback trials than on immediate-feedback trials. For example, in Sassenrath and Yonge's (1969) study, participants attempted to guess the answers to introductory psychology questions and then received feedback of the correct answer for 5 s. For participants who received immediate feedback, the total time spent on each trial was therefore equivalent to the amount of time that it took them to respond, plus 5 s for the feedback presentation. On the other hand, for participants who received 10-s delayed feedback, the total time spent on each trial was equivalent to the amount of time that it took them to respond, plus a 10-s delay, plus a 5-s feedback presentation. After making their responses, therefore, the group that received delayed feedback spent a total of 15 s on each trial, whereas the group that received immediate feedback spent only 5 s on each trial. This design issue was present in the other studies reviewed so far, as well.

The extra time spent on delayed-feedback trials could benefit performance in one of two ways. First, if participants initially guessed the correct answer and then had to wait for 10 s before seeing the answer, they could have spent those 10 s rehearsing or thinking about the answer. Second, even if participants initially guessed the wrong answer, being required to wait 10 s before seeing the feedback may have provided additional opportunities to try to guess the answer. Research on the testing effect (i.e., retrieval practice) has consistently demonstrated that the act of trying to recall information—even if those recall attempts are unsuccessful (e.g., Huelser & Metcalfe, 2012; Kornell, 2014; Kornell, Hays, & Bjork, 2009)—produces significant benefits on later memory for the material (e.g., Carpenter, 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Karpicke & Blunt, 2011; Roediger & Butler, 2011). The additional time afforded by delaying presentation of the correct answer, therefore, could simply be a by-product of the well-known benefits of retrieval practice.

A recent study by Carpenter and Vul (2011) helped rule out the potential confounding effects of rehearsal and retrieval practice on the delay-of-feedback benefit. In this study, participants first saw a series of face–name pairs and then took a cued-recall test in which they had to supply the name when shown each face. In the delayed-feedback condition, participants were given 8 s to try to recall the name, and as soon as they entered a response, they were shown a blank screen for 3 s, followed by presentation of the correct face–name pair for 2 s. In the immediate-feedback condition, participants were also given 8 s to try to recall the name, and then were immediately provided with the correct face–name pair for 5 s. As such, the time spent on each trial in the immediate- and delayed-feedback conditions was 13 s, but the feedback itself was presented for a longer amount of time in the immediate-feedback condition than in the delayed-feedback condition (5 vs. 2 s, respectively). Despite this disadvantage in rehearsal time, delayed feedback still led to greater retention than did immediate feedback on a final test 5 min later.

To examine the second possibility—whether the delay-of-feedback benefit might be influenced by additional time spent trying to retrieve the answer—Carpenter and Vul (2011) included a third condition. In this condition (prolonged test immediate feedback), participants were given 11 s to recall each name and were then provided with the correct answer immediately for 2 s. The total time spent on each of these trials was 13 s, just as in the immediate- and delayed-feedback conditions. However, these trials were unique in that participants had extra time

(11 vs. 8 s) to try to recall the answer before receiving feedback. On the final test, retention of items in the prolonged test immediate feedback condition was no different from retention of items in the immediate-feedback condition, and retention in both of these conditions was inferior to retention in the delayed-feedback condition. Thus, these results demonstrate that the delay-of-feedback benefit does not appear to be driven exclusively by rehearsal time, or by extra time spent trying to retrieve the answer.

What, then, might explain the delay-of-feedback benefit? Carpenter and Vul (2011) proposed that during the delay, participants might actively anticipate the feedback, resulting in a heightened sense of arousal or attention to the feedback when it occurs. This process is less likely to occur when the feedback is immediate, because the instantaneous presence of the answer cuts short any opportunity to engage in this anticipatory process. Consistent with this idea, Carpenter and Vul (Exp. 3) found that the delay-of-feedback benefit was eliminated under conditions in which participants had to perform an attention-demanding distractor task (i.e., counting shapes on the computer screen) during the 3-s delay. When the delay involved a 3-s blank screen with no distractor activity, the usual delay-of-feedback benefit emerged.

Although the study by Carpenter and Vul (2011) provides some theoretical insight into the delay-of-feedback benefit, more studies are needed that can identify the factors underlying this effect. In the present study, we took a new approach to exploring whether the delay-of-feedback benefit is driven by anticipatory processing. Unlike the study by Carpenter and Vul, in which the degree of anticipatory processing was manipulated by task demands, the present study explored the presence of the delay-of-feedback benefit for items that were more versus less likely to be anticipated.

In the present study, participants answered a series of obscure trivia questions (e.g., *What does a deltiologist collect?*). Immediately afterward, participants were asked to rate (on a scale from 1 to 6) how curious they were to know the answer to that question. Feedback of the correct answer (e.g., *Postcards*) was then provided either immediately or after a 4-s delay. Participants' tendencies to anticipate the feedback would presumably occur to a greater degree for items that aroused higher levels of curiosity. If this anticipatory processing drives the delay-of-feedback benefit, then this benefit should be greater for items that are more likely to be anticipated (i.e., those that arouse high levels of curiosity). The key result that would support this hypothesis would be an interaction between curiosity and feedback timing, such that the delay-of-feedback benefit was more likely to occur for items that aroused high versus low levels of curiosity.

Experiment 1

In Experiment 1, we explored whether the delay-of-feedback benefit depends on participants' curiosity to know the answer to a question. Participants were presented with 70 trivia questions, one at a time, and asked to provide their best guess as to the answer. Immediately after typing in their guess, participants rated their curiosity to know the answer to that question using a scale that ranged from 1 (*not at all curious*) to 6 (*very curious*). For each participant, half of the items were randomly assigned to receive feedback immediately, and half after a 4-s delay. A final test over the same questions was then administered after a 5-min retention interval.

Method

Participants A group of 32 undergraduate students at Iowa State University participated in Experiment 1. They were recruited from introductory psychology courses and were given partial course credit upon completing the experiment.

Materials, design, and procedure The 70 trivia questions (see the Appendix) were obtained from a book of obscure trivia facts (Botham, 2006). Participants were first informed that they would be learning a list of trivia facts and that their knowledge of the answers would later be tested. They were then presented with the 70 questions, one at a time, in random order. Each question appeared in the center of the screen, and participants were instructed to "Please type in your best guess of the answer to this question, then press the ENTER key." Thus, participants were encouraged, but not required, to enter a response. After pressing ENTER, the response that participants had typed disappeared from the screen, but the question remained on the screen and participants were asked, "How curious are you to know the answer to this question?" A scale ranging from 1 (*not at all curious*) to 6 (*very curious*) appeared at the bottom of the screen, and participants were asked to enter a number between 1 and 6. They were allowed unlimited time to answer the question and rate their curiosity.

After entering their curiosity rating, feedback of the correct answer was provided directly below the question in one of two ways. For half of the questions, the feedback was displayed for a period of 10 s as soon as the participant entered a curiosity rating (*immediate feedback*). For the other half of the questions, participants were shown the correct answer below the question for 6 s, but after a 4-s delay, during which only the question was displayed on the screen (*delayed feedback*). In this way, the amount of time spent on each trial after entering a curiosity rating was the same for both conditions (i.e., 10 s), although it is worth noting that the feedback itself was presented for a longer amount of time in the immediate than in the delayed condition. The order of presentation and assignment of questions to receive either immediate or delayed feedback was randomly determined for each participant.

After the participants had answered and rated their curiosity for all of the questions, they were asked to type as many U.S. states as they could in 5 min. Participants then completed a final test over the same 70 trivia questions in a new random order. Participants were again allowed unlimited time to answer each question, but on the final test they were not asked to rate their curiosity, nor were they provided with feedback. At completion of the final test, the participants were debriefed and given course credit.

Results and discussion

Scoring All responses were hand scored and given full credit if they were an exact match to the correct answer or contained minor spelling errors. Half credit was given to responses that were considered only partially correct. Two raters who were blind to the conditions scored 25 % of the responses, and the interrater correlations were significant for both the immediate and delayed conditions (rs > .98, ps < .001). The remainder of the scoring was completed by a single rater who was blind to the conditions.

Initial test performance Accuracy on the initial test was low (6 % on average). Errors of omission occurred 13 % of the time across both conditions. Errors of commission consisted of guesses that were appropriate to the questions (e.g., participants typed in a number for those questions asking about a number, a name for those questions asking about a name, etc.), indicating that participants' erroneous guesses appeared to be plausible and consistent with the correct category of response. The average curiosity ratings did not differ between items that received immediate feedback (M = 3.18, SD = 1.13) versus delayed feedback (M = 3.23, SD = 1.15), $t(31) = 0.82$, $p = .42$.

Final test performance Analysis of the final test accuracy was based on the 94 % of items that were answered incorrectly on the initial test. Out of these questions, the proportions answered correctly on the final test were analyzed as a function of both curiosity and feedback timing. The curiosity rating given to each question during the initial test was dichotomized as being either high or low, according to whether that item had received a rating that fell within the upper half (4–6) or the lower half (1–3) of the scale, respectively. Accuracy on the final test was therefore examined across four conditions: (1) high-curiosity items that received immediate feedback, (2) high-curiosity items that received delayed feedback, (3) low-curiosity items that received immediate feedback, and (4) low-curiosity items that received delayed feedback.

Table 1 presents the proportion of answers that were correctly recalled on the final test in each of these four conditions. The following analyses were based on the 25 participants who contributed at least one data point to each of these conditions. A 2 × 2 (Feedback Delay × Curiosity) repeated measures analysis of variance (ANOVA) revealed a significant main effect of curiosity, $F(1, 24) = 14.05$, $p = .001$, $MSE = .007$, $\eta_p^2 = .37$, in that final test accuracy was better for items that were given a high curiosity rating (M = .88, SD = .08) rather than a low curiosity rating (M = .82, SD = .10). No overall main effect of feedback timing emerged, $F(1, 24) = 0.56$, $p = .46$.

A significant Feedback Delay × Curiosity interaction also emerged, $F(1, 24) = 6.42$, $p = .018$, $MSE = .007$, $\eta_p^2 = .21$. For items rated low in curiosity, final test retention did not differ according to whether feedback was immediate or delayed, $t(24) = 0.89$, $p = .38$. However, for items rated high in curiosity, final test retention was significantly higher for items that had received delayed rather than immediate feedback, $t(24) = 2.94$, $p = .007$, $d = 0.59$. This pattern was apparent as well when final test retention was calculated across all six curiosity ratings (see Table 2).

The results of Experiment 1 support the idea that items that were more likely to arouse curiosity were also more likely to benefit from delayed feedback. However, we cannot know whether the results of Experiment 1 were driven by the feedback delay per se, or rather by the *uncertainty* of the delay. Manipulating feedback delay within subjects created conditions under which participants never knew ahead of time on any given trial whether feedback would be provided immediately or after a delay. Studies have shown that sustained attention is enhanced under conditions in which the timing of an upcoming event is uncertain (e.g., Sarter, Givens, & Bruno, 2001). To the extent that the uncertainty of the feedback timing contributed to participants' anticipation of the feedback, the delay-of-feedback benefit in Experiment 1 could have been driven by the uncertainty of the feedback timing and not by the delay itself. The purpose of Experiment 2 was to independently manipulate these two factors in order to evaluate their separate influences on the delay-of-feedback benefit.

Experiment 2

In Experiment 2, participants answered the same trivia questions from Experiment 1 and then rated their curiosity to know the answers using the same scale as before. Unlike in Experiment 1, however, feedback delay was manipulated between subjects and a new condition was added. After making the curiosity rating, one group of participants received feedback immediately (the *immediate group*), another group received feedback that

Table 1 Proportions of answers recalled on the final test as a function of curiosity and feedback timing

	Feedback Condition		
Experiment 1	Immediate	Delayed	
Low curiosity (items rated 1–3)	.835 (.120)	.809 (.136)	
High curiosity (items rated 4–6)	.856 (.106)	.911 (.091)	
Experiment 2	Immediate	Constant Delay	Varied Delay
Low curiosity (items rated 1–3)	.813 (.158)	.816 (.162)	.804 (.172)
High curiosity (items rated 4–6)	.841 (.151)	.859 (.151)	.892 (.112)

Standard deviations are given in parentheses. These proportions are based on participants who contributed at least one item to each of the possible conditions (in Exp. 1, $n = 25$; in Exp. 2, $n = 66$ in the immediate-feedback group, $n = 67$ in the constant-delay group, and $n = 67$ in the varied-delay group). In Experiment 1, the mean numbers of items in the four conditions ranged between 15 and 18. In Experiment 2, the mean numbers of items in the six conditions ranged between 28 and 37.

was delayed by 4 s on every trial (the *constant-delay group*), and the last group received feedback that was delayed by an unpredictable duration—2, 4, or 8 s—on each trial (the *varied-delay group*). Participants in the immediate- and constant-delay groups were informed ahead of time of the exact timing of the feedback they would experience, and received the same feedback timing throughout the experiment. However, those in the varied-delay group, similar to those in Experiment 1, never knew ahead of time exactly when the feedback would occur on any given trial.

If the results of Experiment 1 were driven by the delay of feedback per se, and not by the uncertainty of the delay, then Experiment 2 should replicate the same Feedback Delay × Curiosity ınteraction (showing a larger delay-of-feedback benefit for high-curiosity items than for low-curiosity items) in both the constant- and varied-delay groups. On the other hand, if uncertainty about when the feedback will occur contributes significantly to this finding, then this interaction should be more likely to occur in Experiment 2 for the varied-delay group than for the constant-delay group.

Method

Participants A total of 208 undergraduate students at Iowa State University participated in Experiment 2. They were recruited from introductory psychology courses and were given partial course credit upon completing the experiment. Each participant was randomly assigned to one of three feedback groups: 70 were assigned to the immediate group, 68 to the constant-delay group, and 70 to the varied-delay group.

Design, materials, and procedure Participants learned the same trivia questions from Experiment 1. As in Experiment 1, the participants were first told that they would be learning trivia facts and that they would later be tested over these facts. Each question appeared in the center of the screen, and

Table 2 Proportions of answers recalled on the final test as a function of curiosity rating (1–6) and feedback timing

	Curiosity Rating					
	1	2	3	4	5	6
Experiment 1						
Immediate	.768 (.332)	.788 (.258)	.876 (.154)	.841 (.133)	.893 (.128)	.861 (.247)
Delayed	.794 (.161)	.817 (.216)	.838 (.236)	.906 (.109)	.927 (.105)	.935 (.155)
Experiment 2						
Immediate	.743 (.245)	.818 (.180)	.846 (.174)	.826 (.170)	.815 (.278)	.831 (.237)
Constant	.791 (.207)	.799 (.237)	.843 (.183)	.872 (.160)	.858 (.216)	.919 (.105)
Varied	.815 (.181)	.807 (.172)	.829 (.208)	.884 (.155)	.898 (.172)	.931 (.107)

Standard deviations are given in parentheses. These values are reported for descriptive purposes, because fewer than half of the participants in each experiment used the full range of the scale. In Experiment 1, the mean numbers of items receiving the 1–6 ratings (in that order) in the immediate condition were 8, 4, 7, 5, 5, and 4, and in the delayed condition were 8, 4, 5, 6, 5, and 4. In Experiment 2, the mean numbers of items receiving the 1–6 ratings (in that order) in the immediate condition were 10, 13, 14, 12, 8, and 8; in the constant-delay condition were 11, 11, 12, 11, 7, and 11; and in the varied-delay condition were 11, 13, 13, 12, 7, and 9.

participants were instructed to type in their best guess of the answer and press the ENTER key. Participants were then asked to rate their curiosity to know the correct answer, using the same 1–6 scale as in Experiment 1. They were given unlimited time to answer the questions and rate their curiosity.

Immediately after entering a curiosity rating, feedback was administered according to one of the following between-subjects conditions: (1) immediate, in which feedback was provided immediately after a response was entered; (2) constant delay, in which a 4-s delay was experienced, and then feedback was presented; or (3) varied delay, in which one third of the items were randomly assigned on an individual participant basis to have their answers displayed after a 2-s delay, one third after a 4-s delay, and one third after an 8-s delay. To ensure that equal numbers of items were assigned to receive feedback after 2-s, 4-s, and 8-s delays, one trivia fact from Experiment 1 was chosen at random to be eliminated (i.e., *What was the biggest-selling toy in 1957?*), resulting in 69 total facts that were learned by all participants in Experiment 2. In the varied-delay group, 23 facts were randomly assigned for each participant to receive feedback after 2, 4, or 8 s, and the order of presentation was randomized so that participants never knew ahead of time on any given trial exactly when feedback would be presented. Across all three groups, feedback of the correct answer was displayed for 6 s. Thus, unlike Experiment 1, in which immediate feedback was displayed longer than delayed feedback, in Experiment 2 feedback was displayed for equivalent amounts of time across all conditions. In both the constant-delay and varied-delay groups, only the question remained on the screen during the delay period.

Each of the 69 items was presented in a unique random order for each participant. After participants had answered and received feedback for each item, they completed two unrelated distractor activities lasting a total of 15 min. The first was the same U.S. state-naming task from Experiment 1, in which participants were given 5 min to type as many states as they could. The second part consisted of a 25-letter × 25-letter word search puzzle, below which a list of 35 words was displayed. Participants were told to circle each of the words that they found in the puzzle and then to cross the word out in the list below the puzzle. They were given 10 min to find as many words as possible. None of the words included in the puzzle were answers to any of the trivia questions.

After completing the distractor task, participants took a final test over the same 69 questions, in a new random order. The questions were presented one at a time, and participants were given unlimited time to type in their answers. As in Experiment 1, they did not rate their curiosity, nor did they receive feedback during the final test.

Results and discussion

Scoring Responses were scored according to the same system as in Experiment 1. The interrater correlations for 20 % of the responses were significant across all three groups (rs > .97, ps < .001), so the remainder of the responses were scored by a single rater who was blind to the conditions.

Initial test performance As in Experiment 1, accuracy on the initial test was quite low (7.9 %, on average). Initial test accuracy did not differ across the three conditions, $F(2, 205) = 0.547$, $p = .579$. Errors of omission occurred 11 % of the time in the immediate and varied-delay groups, and 10 % of the time in the constant-delay group. Errors of commission again consisted of guesses that appeared to be plausible and consistent with the correct category of response. The average curiosity ratings did not differ between the immediate-feedback group ($M = 3.32$, $SD = 1.04$), the constant-delay group ($M = 3.37$, $SD = 1.19$), and the varied-delay group ($M = 3.27$, $SD = 1.08$), $F = 0.14$, $p = .87$.

Final test performance As in Experiment 1, the analysis of final test performance included only those items that participants had answered incorrectly on the initial test. Final test performance was again examined as a function of the curiosity ratings that items had received, which were dichotomized as either high (items rated 4–6) or low (items rated 1–3).

Table 1 presents the proportions of answers that were correctly recalled on the final test across the three groups as a function of whether items received high versus low curiosity ratings. To examine performance in the same way as in Experiment 1, we conducted two separate 2 × 2 (Feedback Delay × Curiosity) ANOVAs. The first of these ANOVAs included Delay of Feedback (varied delay vs. immediate) as the between-subjects factor and Curiosity (high vs. low) as the within-subjects factor. The second ANOVA included Delay of Feedback (constant delay vs. immediate) as the between-subjects factor and Curiosity (high vs. low) as the within-subjects factor. Replicating the same interaction from Experiment 1 in both ANOVAs would indicate that the delay of feedback per se was responsible for this effect. On the other hand, replicating the interaction in just the first ANOVA, but not in the second, would indicate that the uncertainty of the feedback delay played a significant role beyond just the delay itself.

Consistent with the latter prediction, a significant Feedback Delay × Curiosity interaction emerged from the first ANOVA,

$F(1, 131) = 7.45$, $p = .007$, $MSE = .008$, $\eta_p^2 = .05$, but not from the second ANOVA, $F(1, 131) = 0.43$, $p = .512$. Both analyses revealed a significant main effect of curiosity, Fs > 9, ps < .002, η_p^2s = .18 and .07, respectively, but no significant overall main effect of feedback timing, Fs < 1. Thus, the same interaction from Experiment 1 was replicated, but only under conditions in which the duration of the feedback delay was varied. As in Experiment 1, this interaction was manifested by a significant delay-of-feedback benefit for items that were rated high in curiosity, $t(131) = 2.24$, $p = .027$, $d = 0.38$, and an absence of this benefit for items that were rated low in curiosity, $t(131) = 0.29$, $p = .770$.

These results replicate the findings from Experiment 1 showing that delayed feedback benefits retention more than does immediate feedback, but only for items that are rated high in curiosity. Furthermore, the results of Experiment 2 show that the delay-of-feedback benefit on high-curiosity items is stronger under conditions in which the timing of the feedback is varied rather than constant. This pattern was apparent as well when final test retention was calculated across all six curiosity ratings (see Table 2). Thus, the delay-of-feedback benefit appears to be influenced by two things: (1) a sense of curiosity to know the answer, and (2) a sense of uncertainty about when that answer will appear.

We now turn to a separate but related question: Does the duration of the delay itself have any effect on learning? In the varied-delay group, items received feedback after either 2, 4, or 8 s. This allowed us to examine final test retention for feedback delays of varying durations. Table 3 presents the proportions of answers that were correctly recalled on the final test as a function of whether they had received high versus low curiosity ratings and whether feedback was received after 2, 4, or 8 s. For items rated low in curiosity, the proportions retained across the three feedback delay durations were similar to the proportion retained by the immediate group. Planned comparisons using independent-samples t tests confirmed the lack of significant differences in final test retention for low-curiosity items retained by the immediate group versus items from the varied-delay group that had received a 2-s, 4-s, or 8-s delay, ts < 1, ps > .50.

For items rated high in curiosity, the proportions retained across the three feedback delay durations were higher than the proportion retained by the immediate group. Planned comparisons using independent-samples t tests confirmed that final test retention was better for items that had received a 2-s feedback delay as compared to immediate feedback, $t(121) = 2.30$, $p = .023$, $d = 0.42$, and for items that had received a 4-s feedback delay as compared to immediate feedback, $t(121) = 2.83$, $p = .005$, $d = 0.51$. Items that had received an 8-s feedback delay were also retained better than those that had received immediate feedback, but this difference was not as strong, $t(121) = 1.97$, $p = .051$, $d = 0.36$. Thus, the delay-of-feedback benefit for high-curiosity items appears to be strongest for items that receive a 4-s delay, suggesting that longer delays of feedback may not necessarily be better. Furthermore, a small but reliable advantage occurred for 4-s delayed feedback in the varied-delay group (90.9 %) as compared to the constant-delay group (85.9 %), $t(122) = 2.06$, $p = .042$, $d = 0.38$. The duration of the feedback delay was the same in both cases, but was less predictable in the varied-delay than in the constant-delay group.

Table 3 Proportions of answers recalled from the varied-delay group on the final test in Experiment 2 as a function of curiosity and feedback delay duration

	Feedback Delay Duration		
	2 s	4 s	8 s
Low curiosity (items rated 1–3)	.790 (.225)	.826 (.157)	.822 (.162)
High curiosity (items rated 4–6)	.900 (.130)	.909 (.109)	.896 (.158)

Standard deviations are given in parentheses. These proportions are based on the 57 participants from the varied-delay group who contributed at least one item to each of these six conditions. The mean numbers of items in the six conditions ranged between 10 and 12.

General discussion

Delaying feedback by brief intervals of time, as compared to providing it immediately, led to better retention of trivia facts that aroused high levels of curiosity. The benefits of delayed feedback over immediate feedback have been observed in other verbal-learning tasks (e.g., Carpenter & Vul, 2011; Sassenrath & Yonge, 1969; Schroth & Lund, 1993), as well as in tasks of complex skill acquisition. For example, Swinnen, Schmidt, Nicholson, and Shapiro (1990) had participants perform a sequence of movements with a manual slide, and then provided them with feedback regarding the speed with which they had executed these movements. One group of participants received this feedback immediately (immediate-feedback group), and another group received it after an 8-s delay (delayed-feedback group). A third group also received feedback after an 8-s delay (estimation group), but these participants were asked to spend the delay trying to estimate their movement time before seeing the feedback. On a later test involving the same task, the estimation group significantly outperformed the immediate-feedback group (the delayed-feedback group also outperformed the immediate-feedback group, but not to a significant degree). Although it is not clear whether this process is the same one that underlies the delay-of-feedback benefit in verbal memory tasks, these results converge with those of the present study by showing that learning benefits by withholding knowledge of the results for a brief period of time following a response.

Unlike many previous studies, we provided only one learning trial and examined retention only for questions that participants answered *incorrectly* on the initial test. These factors could explain why the overall main effect of feedback timing was not stronger in the present study (i.e., only high-curiosity items benefited from delayed feedback). Previous studies on this topic have typically included multiple learning trials over the same items, so the manipulation of feedback timing would apply to some items that had been answered correctly and some that had been answered incorrectly. The extra rehearsal time afforded by delayed feedback following a correct response could be a partial contributor to the larger delay-of-feedback benefit found in previous multitrial experiments. Although Sassenrath and Yonge (1969) observed a small delay-of-feedback benefit in a single-trial experiment, their test questions were constructed from information that participants were learning in their courses (for which they scored over 30 % accurate on the initial test), so that, as in the multitrial experiments, the final test performance consisted of some answers that had initially been correct.

The delay-of-feedback benefit in the present study was observed for items that aroused high levels of curiosity, but not for items that aroused low levels of curiosity. Given that participants would seem more likely to anticipate the answer to a question about which they are highly curious, this pattern of results seems consistent with the idea that the delay-of-feedback benefit is driven by active anticipatory processing that occurs during the delay.

What might be the nature of this anticipatory processing? One possibility is that participants engage in reflection on their own response during the delay. Previous studies have shown, on the basis of subjective reports, that participants tend to continue thinking about their own response after answering a question (e.g., Brackbill & Kappy, 1962). For questions that arouse high curiosity, it seems plausible that this tendency would be stronger, due perhaps to greater interest in the question or greater motivation to know the answer. This response-based processing is more likely to occur when feedback is delayed rather than immediate, because the former allows more time for such processing to take place. Such processing may benefit later retention of the correct response by enhancing source memory for the original (erroneous) response and facilitating its discrimination from the correct response, or perhaps by serving as mediating information that helps strengthen the link between the question and the correct answer (see, e.g., Brackbill & Kappy, 1962; Parlow & Berlyne, 1971).

Indeed, recent research has shown that good memory for initial error responses could play a role in successful error correction. Long-term retention of correct responses is greater when participants remember their own original errors than when they do not (e.g., Butler, Fazio, & Marsh, 2011; Peeck, van den Bosch, & Kreupeling, 1985), suggesting—somewhat counterintuitively—that effective error correction may depend on good memory for one's initial error. Thus, the nature of the anticipatory processing that takes place during a delay of feedback could be driven (at least partially) by the extra time afforded by the delay to engage in processing of one's own response. Such reasoning could explain the tendency for the delay-of-feedback benefit to be stronger under conditions in which participants are encouraged to think about their own response during the delay (e.g., Swinnen et al., 1990), and weaker under conditions in which participants' processing of their own responses is hindered by performing a secondary task during the delay (Carpenter & Vul, 2011).

During the delay, participants may also engage in enhanced processing of the question itself, continue to puzzle over what the answer might be, and attempt to resolve the sense of uncertainty that arises from not knowing the answer. Consistent with this notion, past research on the effects of "effort after meaning" has found that participants' memory for ambiguous sentences (e.g., "The notes were sour because the seam split") benefits by inserting a brief delay of several seconds before the presentation of a disambiguating word (e.g., "bagpipe") that gives meaning to the sentence (e.g., Auble & Franks, 1978; see also Zaromb, Karpicke, & Roediger, 2010). Along similar lines, Kronlund and Whittlesea (2006) found that sentences were remembered better if the highly constraining sentence stem (e.g., "She swept the kitchen floor with a . . .") was followed by a pause of a few seconds, versus no pause, before presenting the final word ("broom"). Interestingly, Zaromb and Roediger (2009) found that this effect only occurred in a within-subjects design in which, on a random half of trials, the disambiguating word was either embedded in the sentence or delayed by several seconds after presentation of the sentence. When participants were given only sentences in which the disambiguating word was either embedded or delayed, the delay was not beneficial. This finding parallels that of Experiment 2, showing that the delay-of-feedback benefit only occurred when the duration of the feedback delay was varied and unpredictable. It is possible that the ambiguity that arises from not knowing the answer could be enhanced when the feedback delay is varied rather than constant, because the former situation creates a sense of uncertainty not only about the correct answer itself, but also about *when* the uncertainty will be resolved. Although these hypotheses have not been directly applied to any of the known research on feedback timing, they provide some promising ideas that can be directly explored in future studies.

In the present study, the processing that takes place during the delay could depend on the factor(s) that drive participants' curiosity. One factor influencing curiosity could be some degree of partial knowledge that participants have about the answer to a given question. For example, even if they could not recall what a deltiologist collects, it is possible that participants have some familiarity with this word (perhaps having heard it before but not knowing what it means), and this sense of familiarity could drive their curiosity to see the answer. Curiosity judgments could also reflect confidence: Participants may be more curious about answers that they feel they have gotten correct or that they feel they *should* know. If curiosity is positively related to confidence, then the memory advantage for high-curiosity over low-curiosity items could represent a type of hypercorrection effect (see, e.g., Butterfield & Metcalfe, 2001, 2006), particularly if the correct answer deviates from one's expectations (Fazio & Marsh, 2009).

In addition to studies that manipulate feedback delay over brief time periods of a few seconds, a parallel body of research has explored the effects of feedback delayed by much longer time intervals of at least 24 h. These studies show that learning benefits from feedback that is delayed by 24 h after taking a test, as compared to feedback that is provided on the same day as the test (e.g., Butler, Karpicke, & Roediger, 2007; Kulhavy & Anderson, 1972; Kulik & Kulik, 1988; Metcalfe, Kornell, & Finn, 2009; Surber & Anderson, 1975). In one recent study (Mullet, Butler, Verdin, von Borries, & Marsh, 2014), students in a college-level engineering course completed homework assignments and were then given access to the answers either immediately after completing the assignments or one week later. On course exams requiring transfer of the principles learned through homework, students who had received one-week delayed feedback significantly outperformed their classmates who had received immediate feedback. It has been proposed that these longer feedback delays may be beneficial because the delay could provide distributed practice (i.e., a spacing effect) for those items that students get correct (e.g., Butler et al., 2007). Another theoretical account—the *interference perseveration hypothesis* proposed by Kulhavy and Anderson (1972)—proposes that the delay allows time for students to forget their initial error responses, which serves to reduce interference between the error and the correct response when it is presented. More recent data, however, have shown that students' initial errors do not appear to interfere with learning the correct response (e.g., Butler & Roediger, 2008), and may actually be beneficial for error correction (Butler et al., 2011). Although the mechanisms underlying the benefits of delayed feedback may be different for feedback that is delayed by very brief intervals (e.g., 4 s) versus longer intervals (e.g., 24 h), it is encouraging to note that, from a practical perspective, students' learning does not appear to be harmed by delaying feedback of the correct answers after taking a test.

Both of the present experiments revealed significant main effects of curiosity, in that high-curiosity items were retained better than low-curiosity items. Although we assume that participants' curiosity ratings reflected their intrinsic sense of curiosity to know the answers, it is possible that making this judgment on an item-by-item basis could have directed participants' attention to their own curiosity and affected responding in ways that would not have occurred had they not been required to make an explicit judgment for each item. Previous research has shown, however, that curiosity and memory are positively related even when a judgment of curiosity does not immediately precede the answer to a given item. For example, Berlyne (1954) gave participants an initial test over a series of questions about animals, and participants were told to mark the questions for which they were curious to know the answers. They then read information about each animal that provided answers to these initial questions. On a final test over the same questions, participants scored higher on questions about which they had been curious to know the answers, as compared to questions about which they had not been curious to know the answers. Thus, curiosity and memory were positively related even when there was no immediate link between making the curiosity judgment for a given item and then processing the information to which it pertained. Other studies as well have reported a positive association between curiosity and memory for materials such as general knowledge questions (Kang et al., 2009), authors of famous quotations (Berlyne, 1966), short stories (Maw & Maw, 1961), and products in advertisements (Bull & Dizney, 1973).

In conclusion, these results may have important implications for teaching and instruction. In particular, they suggest that students' retention of material might benefit by inserting brief pauses after asking questions. This could be easily accomplished in computer-assisted instruction by simply delaying the presentation of the correct answer by a few seconds. In classroom instruction, teachers might also find it beneficial to pause for brief periods of time after posing questions to students. Research on teacher "wait time" supports this notion, showing that brief pauses of a few seconds following questions can positively influence students' participation in class, as well as their memory for course material (e.g., Tobin, 1987). In light of the present findings, future research should explore whether techniques designed to arouse students' curiosity contribute further to the beneficial effects of teacher wait time.

Author note This study was conducted as part of the requirements for the degree of Master of Science by the first author. Portions of this study were presented at the annual meeting of the Psychonomic Society (Seattle, WA, November 2011). We thank Ashley Adams, Syamim Hasim, Lauren Miller, Courtney Tapp, and Andrew Woods for their assistance with data collection and scoring.

Appendix

Table 4 List of trivia questions used in the present study

	Question	Answer
1.	What does a deltiologist collect?	Postcards
2.	Who appeared on the cover of *Life* magazine more than anyone else?	Elizabeth Taylor
3.	In what country did doughnuts originate?	Holland
4.	What do phobatrivaphobics fear?	Trivia about phobias
5.	What was the top-rated television series from 1957–1961?	*Gunsmoke*
6.	Where did the Beatles perform their first U.S. concert?	Carnegie Hall
7.	How many U.S. presidents were the only child in their families?	0
8.	What was Adolf Hitler's favorite movie?	*King Kong*
9.	How long did Leonardo da Vinci spend painting the Mona Lisa's lips?	12 years
10.	What do dendrologists study?	Trees
11.	What is the largest object ever found in the Los Angeles sewer system?	Motorcycle
12.	What is the most preferred reading material for the bathroom?	*Reader's Digest*
13.	What does a librocubicularist do?	Reads in bed
14.	What is the world's most read comic strip?	*Peanuts*
15.	What is a prestidigitator?	Magician
16.	What is the most popular ice cream flavor?	Vanilla
17.	Coca-Cola was originally what color?	Green
18.	What is the average number of houses a person looks at before buying one?	8
19.	What was the first novel written on a typewriter?	*Tom Sawyer*
20.	Who was the first female monster to appear on the big screen?	Bride of Frankenstein
21.	Which country offered Albert Einstein its presidency in 1952?	Israel
22.	In which season do most burglaries occur?	Winter
23.	Where did the yo-yo originate?	The Philippines
24.	Thurl Ravenscroft was the voice of what cartoon character?	Tony the Tiger
25.	What size of shoe did Robert E. Lee wear?	4.5
26.	What animal cannot contract or carry the rabies virus?	Squirrels
27.	The ruins of Troy are located in what country?	Turkey
28.	How many spaces are on a Scrabble board?	225
29.	What is the most common name for a goldfish?	Jaws
30.	What was the biggest-selling toy in 1957?	The hula hoop
31.	What song is sung most often in America?	"Happy Birthday"
32.	What letter does not appear on the periodic table of elements?	J
33.	What was the first domesticated animal?	Goose
34.	What year in a marriage is the leather anniversary?	Third
35.	What is the smallest unit of time?	The yoctosecond
36.	Which U.S. president wrote 37 books?	Theodore Roosevelt
37.	What is the oldest word in the English language?	Town
38.	What does Pokémon stand for?	Pocket Monster
39.	What color is a grasshopper's blood?	White
40.	Where did voodoo originate?	Haiti
41.	What is the oldest known vegetable?	Pea
42.	Who was the sun god of ancient Egypt?	Ra
43.	What is the top-grossing Disney movie of all time?	*The Lion King*
44.	What is the longest one-syllable word in the English language?	Screeched
45.	What does the word calendar mean?	To call out
46.	What is a group of kangaroos called?	Mob
47.	What is the most common name in the world?	Mohammed
48.	To which fruit family does an almond belong?	Peach
49.	What is the hottest chili in the world?	Habanero
50.	What are the plastic things on the end of shoelaces called?	Aglets
51.	The world's smallest painting is on the surface of what?	Grain of corn
52.	What is the side of a hammer called?	Cheek
53.	Which zoo has the largest collection of animals in the world?	San Diego Zoo
54.	Which country is the largest exporter of frogs' legs?	Japan
55.	What does camera shutter speed "B" stand for?	Bulb
56.	What is IBM's motto?	Think
57.	Who coined the word "nerd"?	Dr. Seuss
58.	The most presidents have been born in which state?	Virginia
59.	How many muscles does a caterpillar have?	4,000
60.	Which U.S. president was once a male model?	Gerald Ford
61.	What does a horologist measure?	Time
62.	What is considered to be the sister language of English?	German
63.	Which native Mexican group went to battle with wooden swords so as not to kill their enemies?	Toltecs
64.	What is a baby bat called?	Pup
65.	In what year was the Golden Gate Bridge first opened?	1937
66.	What do the Olympic rings represent?	The continents
67.	What does *karaoke* mean in Japanese?	Empty orchestra
68.	Which famous composer wrote "Twinkle Twinkle Little Star" at the age of five?	Mozart
69.	What comedic actor was voted the least likely to succeed in high school?	Robin Williams
70.	Which U.S. city sells more popcorn than anywhere else in the country?	Dallas

References

Auble, P. M., & Franks, J. J. (1978). The effects of effort toward comprehension on recall. *Memory & Cognition, 6,* 20–25. doi:10.3758/BF03197424

Berlyne, D. E. (1954). An experimental study of human curiosity. *British Journal of Psychology, 45,* 256–265.

Berlyne, D. E. (1966). Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology, 57,* 128–132.

Botham, N. (2006). *The book of useless information*. New York, NY: Penguin.

Brackbill, Y. (1964). The impairment of learning under immediate reinforcement. *Journal of Experimental Child Psychology, 1,* 199–207.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962a). Delayed-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology, 55,* 947–952.

Brackbill, Y., Isaacs, R. B., & Smelkinson, N. (1962b). Delay of reinforcement and the retention of unfamiliar, meaningless material. *Psychological Reports, 11,* 553–554.

Brackbill, Y., & Kappy, M. S. (1962). Delay of reinforcement and retention. *Journal of Comparative and Physiological Psychology, 55,* 14–18.

Bull, S. G., & Dizney, H. F. (1973). Epistemic-curiosity-arousing prequestions: Their effect on long-term retention. *Journal of Educational Psychology, 65,* 45–49.

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review, 18,* 1238–1244. doi:10.3758/s13423-011-0173-y

Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13,* 273–281. doi:10.1037/1076-898X.13.4.273

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. doi:10.3758/MC.36.3.604

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1491–1494. doi:10.1037/0278-7393.27.6.1491

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

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21,* 279–283. doi:10.1177/0963721412452728

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review, 19,* 81–86. doi:10.3758/s13423-011-0185-7

Carpenter, S. K., & Vul, E. (2011). Delaying feedback by three seconds benefits retention of face–name pairs: The role of active anticipatory processing. *Memory & Cognition, 37,* 1211–1221. doi:10.3758/s13421-011-0092-1

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58.

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

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

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & Cognition, 38,* 951–961. doi:10.3758/MC.38.7.951

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40,* 514–527. doi:10.3758/s13421-011-0167-z

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20,* 963–973. doi:10.1111/j.1467-9280.2009.02402.x

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103,* 48–59.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. doi:10.1126/science.1199327

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 106–114.

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

Kronlund, A., & Whittlesea, B. W. A. (2006). Remembering after a perception of discrepancy: Out with the old, in with the two. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1174–1184. doi:10.1037/0278-7393.32.5.1174

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63,* 505–512.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97.

Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology, 79,* 320–322.

Maw, W. H., & Maw, E. W. (1961). Information recognition by children with high and low curiosity. *Educational Research Bulletin, 40,* 197–224.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37,* 1077–1087. doi:10.3758/MC.37.8.1077

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*. doi:10.1016/j.jarmac.2014.05.001

Parlow, J., & Berlyne, D. E. (1971). The effect of prior guessing on incidental learning of verbal associations. *Journal of Structural Learning, 2,* 55–65.

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. doi:10.1037/0278-7393.31.1.3

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.

Perin, C. T. (1943). A quantitative investigation of the delay of reinforcement gradient. *Journal of Experimental Psychology, 32,* 37–51.

Pubols, B. H. (1958). Delay of reinforcement, response perseveration, and discrimination reversal. *Journal of Experimental Psychology, 56,* 32–40.

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

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews, 35,* 146–160.

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.

Schroth, M. L. (1992). The effects of delay of feedback on a delayed concept formation transfer task. *Contemporary Educational Psychology, 17,* 78–82.

Schroth, M. L., & Lund, E. (1993). Role of delay of feedback on subsequent pattern recognition transfer tasks. *Contemporary Educational Psychology, 18,* 15–22.

Sturges, P. T., Sarafino, E. P., & Donaldson, P. L. (1968). Delay-retention effect and informative feedback. *Journal of Experimental Psychology, 78,* 357–358.

Surber, J. R., & Anderson, R. C. (1975). Delay-retention effect in natural classroom settings. *Journal of Educational Psychology, 67,* 170–173.

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. doi:10.1037/0278-7393.16.4.706

Tarpy, R. M., & Sawabini, F. L. (1974). Reinforcement delay: A selective review of the last decade. *Psychological Bulletin, 81,* 984–997.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57,* 69–95.

Zaromb, F. M., Karpicke, J. D., & Roediger, H. L., III. (2010). Comprehension as a basis for metacognitive judgments: Effects of effort after meaning on recall and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 552–557. doi:10.1037/a0018277

Zaromb, F. M., & Roediger, H. L., III. (2009). The effects of "effort after meaning" on recall: Differences in within- and between-subjects designs. *Memory & Cognition, 37,* 447–463. doi:10.3758/MC.37.4.447