Journal of Applied Research in Memory and Cognition 3 (2014) 222–229

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately

Hillary G. Mullet[a,*], Andrew C. Butler[a], Berenice Verdin[b], Ricardo von Borries[b], Elizabeth J. Marsh[a]

[a] Duke University, United States
[b] University of Texas at El Paso, United States

ARTICLE INFO

Article history:
Received 5 December 2013
Received in revised form 28 April 2014
Accepted 6 May 2014
Available online 14 May 2014

Keywords:
Immediate feedback
Delayed feedback
Transfer
Learning
Classroom

ABSTRACT

Educators and researchers who study human learning often assume that feedback is most effective when given immediately. However, a growing body of research has challenged this assumption by demonstrating that delaying feedback can facilitate learning. Advocates for immediate feedback have questioned the generalizability of this finding, suggesting that such effects only occur in highly controlled laboratory settings. We report a pair of experiments in which the timing of feedback was manipulated in an upper-level college engineering course. Students practiced applying their knowledge of complex engineering concepts on weekly homework assignments, and then received feedback either immediately after the assignment deadline or 1 week later. When students received delayed feedback, they performed better on subsequent course exams that contained new problems about the same concepts. Although delayed feedback produced superior transfer of knowledge, students reported that they benefited most from immediate feedback, revealing a metacognitive disconnect between actual and perceived effectiveness.

> "In many cases – for example, when papers are taken home to be corrected – as much as 24 hours may intervene [before students receive feedback]. It is surprising that this system has any effect whatsoever." – B.F. Skinner (1954, p. 91)

Although Skinner wrote this statement 60 years ago, the assumption that delaying feedback impairs learning remains popular among researchers and educators today. The practical recommendation from most reviews of the feedback literature is that feedback should be given as soon as possible (e.g., Azevedo & Bernard, 1995; Hattie & Timpersley, 2007; Kulik & Kulik, 1988; Mory, 2004). Likewise, promotional materials for educational products such as classroom response systems (e.g., "Pedagogy in Action," 2013), online courses (e.g., Coursera; "Pedagogical Foundations," 2013), and testing tools (e.g., The Immediate Feedback Assessment Technique; Epstein Educational Enterprises, 2013) emphasize the importance of providing learners with feedback immediately after a response. The primary purpose of the present research was to examine the assumption that delaying feedback is harmful to learning. We conducted two experiments in which we manipulated the timing of feedback on homework assignments in a college course. We also surveyed students about their experience in the course in order to explore the degree of correspondence between the actual and perceived effectiveness of delayed feedback.

1. Background

The assumption that feedback must be given immediately in order to be maximally effective derives from the behaviorist approach to learning in which feedback was conceptualized in terms of reinforcement and punishment (e.g., Hull, 1943; Skinner, 1938; Thorndike, 1932). In operant learning paradigms, researchers used reinforcement and punishment to modify voluntary behavior (e.g., using food pellets to shape a desired key-pressing behavior in pigeons). One of the core findings that emerged from such studies was that the response and subsequent feedback had to be paired closely in time in order for the animal to perceive the contingency for learning (for review see Renner, 1964). In fact, delaying feedback by even a few seconds dramatically decreased its effectiveness (e.g., Saltzman, 1951). The successful use of operant conditioning to facilitate learning inspired behaviorists and

* Corresponding author at: Psychology & Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086, United States. Tel.: +1 919 681 1321; fax: +1 919 660 5726.
E-mail addresses: hillary.mullet@duke.edu, hgm8@duke.edu (H.G. Mullet).

http://dx.doi.org/10.1016/j.jarmac.2014.05.001
2211-3681/

educational psychologists to apply these principles to the classroom (e.g., Holland & Skinner, 1961; Jensen, 1949; Pressey, 1963; Skinner, 1958; Stephens, 1960). In the process, the idea that students benefit most from immediate feedback took root and flourished in education.

Beginning in 1960s, new findings from laboratory studies began to challenge the assumption that delaying feedback hindered learning (e.g., Atkinson, 1969; Battig & Brackett, 1961; Brackbill & Kappy, 1962). As the cognitive revolution unfolded, the mechanistic notion of feedback as a motivator of behavior was replaced by the idea that the function of feedback is to provide information to the learner (for review see Kulhavy, 1977). This re-conceptualization of feedback as information spurred further research that demonstrated that delayed feedback can be more effective in promoting long-term retention than immediate feedback (e.g., Anderson et al., 1972; Butler & Roediger, 2008; Carpenter & Vul, 2011; Metcalfe, Kornell, & Finn, 2009; Sassenrath & Yonge, 1968; Smith & Kimball, 2010). For example, Butler, Karpicke, and Roediger (2007, Experiment 2) had participants read a set of expository passages and then take a test with feedback provided either immediately after each response or after a delay of 24 h. On a final retention test 1 week later, participants performed better on items for which they had received delayed feedback.

Despite ample evidence for the mnemonic benefits of delaying feedback, reviewers of the feedback literature have largely dismissed these findings as occurring under tightly controlled conditions that do not generalize to the classroom (e.g., Azevedo & Bernard, 1995; Kulik & Kulik, 1988; Mory, 2004). For example, Kulik and Kulik (1988) conducted a meta-analysis of studies that have manipulated the timing of feedback. They found that delayed feedback is often superior to immediate feedback in the laboratory, but classroom studies consistently show the opposite pattern. Based on these results, they made the following recommendation to educators: “The message from the practical evaluations therefore seems fairly clear. Teachers who want their quizzes to help students learn should try to arrange conditions so that students receive feedback as quickly as possible after they answer quiz questions” (p. 93).

A closer reading of the timing of feedback literature, especially the studies included in the Kulik and Kulik (1988) meta-analysis, suggests that this conclusion is premature. Many of the studies that show an advantage of immediate over delayed feedback contain one or more serious methodological flaws that undermine their conclusions (Angell, 1949; Little, 1934; Pressey, 1950; Sullivan, Schutz, & Baker, 1971). For example, one study confounded the timing of feedback manipulation by presenting feedback both immediately and after a delay for students in the immediate-feedback group, whereas the delayed-feedback group only saw the feedback once (Paige, 1966). In other studies, the primary dependent measure was speed of acquisition (i.e., how many trials participants took to learn the correct responses) rather than retention (Landsman & Turkewitz, 1962; Markowitz & Renner, 1966; Saltzman, 1951; Sturges, Sarafino, & Donaldson, 1968). Due to the focus on acquisition, these studies cannot speak to the question of how delaying feedback affects subsequent memory for learned information; indeed, factors that increase the rate of initial learning often fail to enhance long-term retention (Bjork & Bjork, 1992). In fact, we know of only one methodologically sound classroom study that shows an advantage of immediate feedback on a delayed final test (White, 1968). In an educational psychology course, White (1968) manipulated the timing of feedback after eight short unit tests given throughout the semester. He found that providing feedback immediately after the unit tests (either after each item or after the entire test) produced superior transfer to new questions on the final exam relative to delaying feedback until three days after the initial tests.

2. The present research

Based on the preceding analysis of the timing of feedback literature, there is a clear need for further research to investigate how the timing of feedback affects learning in the classroom. To this end, we conducted two experiments in successive semesters in an upper-level undergraduate engineering course (“Continuous Time Signals and Systems”) at University of Texas at El Paso (UTEP). The course covers signals, systems, and transforms, and examines both the theoretical underpinnings and practical applications of these concepts. The instructor gave two lectures each week and assigned students additional materials for viewing outside the classroom (e.g., texts, videos, and simulations). Students completed the weekly homework assignments outside of class through OpenStax Tutor, an e-learning system designed to allow experimentation and the implementation of cognitive science principles to benefit student learning (Butler, Marsh, Slavinsky, & Baraniuk, 2014). Feedback on the homework problems was released to students either immediately after the assignment deadline or 1 week after the deadline. Importantly, the feedback message was identical for all students regardless of the timing of feedback. Students were required to view the feedback in order to receive credit for the assignment in Experiment 1, and this requirement was manipulated in Experiment 2.

Student learning from the feedback was assessed on course exams, which required students to transfer their knowledge of the concepts covered in the homework assignments to novel problems. In addition to assessing learning, we also surveyed students about their perceptions of the relative effectiveness of immediate and delayed feedback as well as their personal preferences on the optimal timing of feedback. Importantly, the surveys were conducted at the end of the semester so that students’ responses could be informed by their experiences throughout in the course. In addition, the timing of feedback was manipulated between-students in Experiment 1 and within-students in Experiment 2, which allowed us to compare the perceptions of students who experienced only immediate or delayed feedback in the course with that of students who experienced both timing conditions. The protocols for both studies were approved by the Institutional Review Board at UTEP.

3. Experiment 1

In Experiment 1, we manipulated the timing of feedback across two sections of the course. Students in one section received feedback immediately after the assignment deadline, whereas students in the other section received it 1 week after the deadline.

3.1. Method

Participants. Twenty-six students consented to participate in the experiment by releasing their data from the course for research purposes.

Design. The experiment used a single factor (timing of feedback: immediate, delayed), between-subjects design.

Materials. The materials included 8 topics (e.g., Laplace transforms, convolution, Fourier transforms, etc.) for which there was a corresponding homework assignment (3 additional topics were excluded from the analyses because they were covered in the week immediately preceding an exam, thus making it impossible to implement the timing of feedback manipulation for the corresponding homework). Each topic contained approximately 3–6 core concepts that were covered in lecture during the week prior to the homework assignment. The homework assignments contained 10–14 practice problems related to the core concepts for a given topic. Each practice problem required the application of a

Find the unilateral Laplace transform $X(s)$ of signal $x(t) = \sin(2\pi t)u(t)$, where $u(t)$ is the unit step function.

a) $X(s) = \frac{2\pi}{s^2 + 4\pi^2}$

b) $X(s) = \frac{s}{s^2 - j2\pi}$

c) $X(s) = \frac{2\pi}{s^2 - 2\pi}$

d) $X(s) = \frac{s2\pi}{s^2 - 4\pi^2}$

e) $X(s) = \frac{2\pi}{s^2 - 4\pi^2}$

Detailed Solution

Using the Laplace transform integral, we have

$$X(s) = \int_{0^-}^{\infty} x(t)e^{-st}dt$$

$$= \int_{0^-}^{\infty} \sin(2\pi t)e^{-st}dt$$

From the trigonometric identity $\sin(2\pi t) = \frac{e^{j2\pi t} - e^{-j2\pi t}}{2j}$,

we can write

$$= \int_{0^-}^{\infty} \frac{e^{j2\pi}e^{-st}}{2j}dt - \int_{0^-}^{\infty} \frac{e^{-j2\pi}e^{-st}}{2j}dt$$

$$= \frac{1}{2j}\frac{e^{-(s-j2\pi)t}}{-(s-j2\pi)}\Big|_{0^-}^{\infty} - \frac{1}{2j}\frac{e^{-(s+j2\pi)t}}{-(s+j2\pi)}\Big|_{0^-}^{\infty}$$

Assuming $\mathrm{Re}[s] > 0$, then we have

$$X(s) = \frac{1}{2j}\left[\frac{1}{(s-j2\pi)} - \frac{1}{(s+j2\pi)}\right]$$

$$= \frac{1}{2j}\frac{2j2\pi}{(s-j2\pi)(s+j2\pi)}$$

$$= \frac{2\pi}{(s^2+4\pi^2)}$$

Fig. 1. A sample problem from the experiment. At first, the problem was presented in OpenStax Tutor without any multiple-choice alternatives. After the students worked the problem and entered a free-form response, they were presented with the alternatives and selected a response. Finally, they were notified that feedback was available either immediately after the deadline or 1-week after the deadline. The feedback screen displayed both the problem and the detailed solution.

concept to determine the solution (see Fig. 1 for an example problem). Another set of problems was used for the course exams so that no problems were repeated from the homework. Like the homework problems, each exam problem required the application of a concept. The exams contained 5–10 problems with approximately equal coverage of the various topics from the preceding weeks (the exams did not contain a problem for every concept due to time limitations).

Procedure. Students completed homework assignments on a weekly basis. They had 1 week to work the problems and submit their solutions via OpenStax Tutor (the online homework system; www.openstaxtutor.org). Entering their solution to each problem involved two steps. First, they either entered their solution in a free-form text box or uploaded it into the system as an image file. Second, after submitting the free-form response, they were shown a set of multiple-choice alternatives and asked to select the correct answer. The purpose of this two-step response procedure was to give students practice with generating a solution on their own (i.e., without the potential aid of seeing the multiple-choice alternatives), while also enabling the automated scoring of their multiple-choice responses. The initial free-form responses were not graded.

Feedback on the homework problems was released to students either immediately after the assignment deadline or 1 week after the deadline. Once it had been released, students received an email message indicating that the feedback was available for viewing. In order to view the feedback for a given problem, they clicked on a unique link that opened the feedback screen for that problem. Feedback consisted of a re-presentation of the problem, the correct multiple-choice alternative, and a detailed solution (see Fig. 1). Students were required to view the feedback for each problem in order to receive credit for completing the assignment.

Students took four exams over the course of the semester – three unit exams and one final exam. The exams consisted of new problems about the material that was learned through the homework assignments. All exams were closed book and administered in the classroom using the OpenStax Tutor system. Students used the two-step answering procedure to enter their solutions to the problems. They were required to work individually on all of the course activities that were included in the experiment (homework assignments, feedback viewing, and exams).

3.2. Results

All results were significant at the .05 level unless otherwise stated. Cohen's d is the measure of effect size reported for all t-test analyses and eta-squared is the measure of effect size reported for all ANOVA analyses. Two criteria had to be met for a student's data to be included in the analyses: (1) the student had not taken the course before, and (2) the student had to have completed at least half of the homework assignments (4 or more of 8) and course exams (2 or more of 4). Two students were retaking the course and thus their data were excluded. All of the remaining 24 students met the criteria for inclusion in the analyses.

In addition, data from individual homework assignments and the corresponding exam problems were excluded from the analyses on a student-by-student basis according to the following criteria: (1) the student completed less than half of the problems for the homework assignment; (2) the student completed the assignment after the deadline and within 24 h of the corresponding exam or after the exam. Applying these criteria resulted in the exclusion of the data related to 3% (6 out of 192) of homework assignments.

Homework performance. The proportion of correct multiple-choice responses on the homework assignments was approximately equal between students who received immediate feedback and students who received delayed feedback [0.83 vs. 0.81; $t < 1$]. This result was expected because their responses were made before the timing of feedback manipulation was introduced; nevertheless, it is important because it indicates that any subsequent differences in exam performance were not due to pre-existing differences between the students in the two sections (e.g., inherent ability or learning that occurred before the homework assignments).

Feedback viewing. Overall, students viewed the feedback (i.e., loaded the feedback screen) for 98% of all problems; thus, even though feedback viewing was required, students occasionally forgot or chose not to view it. Students who received feedback immediately after the deadline viewed the feedback for a significantly greater percentage of the problems than those who received delayed feedback [99% vs. 94%; $t(7) = 3.04$, SED = 1.75, $p = 0.02$, $d = 1.87$].[1] In the immediate feedback condition, an average of 4.1 days elapsed between when learners submitted the homework problems and when they viewed the feedback; in the delayed feedback condition, the average lag between homework completion and feedback was 11.6 days. Thus, students did not always view the feedback at precisely the time that it became available; however, as intended, the lag between homework completion and feedback viewing was approximately 1 week longer in the delayed feedback condition than the immediate feedback condition.

On average, students viewed the feedback for each problem 1.54 times (excluding the small percentage of problems for which students did not view feedback at all). Students who received immediate feedback viewed the feedback for a given problem a greater number of times than students who received delayed feedback, but this difference was not significant [1.65 vs. 1.32; $t(22) = 1.66$, SED = 0.20, $p = 0.11$, $d = 0.72$].

Exam performance. On the exams, students who received delayed feedback on the homework assignments answered a greater proportion of multiple-choice questions correctly than did students who received immediate feedback [0.92 vs. 0.84; $t(22) = 2.16$, SED = 0.04, $p = 0.04$, $d = 0.75$] (see footnote 1).

Survey. Nineteen of the 24 students completed the final survey. The overwhelming majority of the students who received immediate feedback reported that they liked the timing of the feedback a lot (82%) and they thought that their performance in the course had benefited a lot from it (73%). In contrast, the majority of students who received delayed feedback reported that they did not like it (57%) and that their learning either did not benefit from it (43%) or was hurt by it (14%). See the Appendix for the full results from the survey.

3.3. Discussion

In contrast to earlier classroom studies that found a superiority of immediate feedback (see Kulik & Kulik, 1988), our experiment showed that delaying feedback by 1 week benefited student performance on course exams. How can these disparate findings be reconciled? One possible explanation is that we made sure that the feedback message and procedure for viewing it was identical regardless of when it was delivered to students. As explained above, many classroom studies have confounded the timing of feedback with other variables, such as the type of feedback (see Butler et al., 2007). One common confound is different requirements for the processing of immediate and delayed feedback (e.g., Pressey, 1950). Immediate feedback is often given in a manner that requires processing of every item (e.g., going over the test in class right after completing it), but delayed feedback is given such that processing is effectively optional (e.g., handing back the marked up exam 1 week later and allowing students to process it on their own time). In contrast, most laboratories studies of feedback timing attempt to ensure that feedback is fully processed regardless of the timing. In Experiment 1, the requirement to view feedback for credit likely motivated students to view the feedback regardless of when it was delivered. However, if students are generally more motivated to view the feedback when it is delivered immediately, then it could explain why immediate feedback produced better learning in previous classroom studies that did not require the processing of delayed feedback. Indeed, none of these studies measured whether students actually viewed the delayed feedback.

Although delaying feedback benefited learning, the end-of-semester survey showed that students largely failed to appreciate these benefits. This finding is consistent with other research showing that learners often have poor insight into the conditions that foster the best long-term retention (Bjork, Dunlosky, & Kornell, 2013). For example, Kornell (2009) found that 90% of participants benefited from a spaced study schedule (i.e., studying that was distributed over time, rather than massed all at one time), but 72% of those students believed that massing was more effective. Such metacognitive illusions are especially likely to arise when the more effective strategy produces initial learning conditions that learners perceive as less fluent and/or more difficult, which is likely to be true with processing delayed feedback.

4. Experiment 2

One goal of Experiment 2 was replicate the key findings from Experiment 1. In order to examine the generalizability of the effect of delaying feedback, we switched from a between-subjects to a within-subjects design. Since students in Experiment 2 experienced both immediate and delayed feedback within the same course, we could assess whether the ability to directly compare these two experiences changed their perceptions of delayed feedback (i.e., relative to Experiment 1). One possible outcome was that students' metacognitive judgments would align more closely with their actual course performance (i.e., they would appreciate the benefits of delayed feedback for their learning).

A second goal of Experiment 2 was to examine one potential explanation for the discrepancy between our study and previous classroom studies regarding the optimal timing of feedback. As explained above, students were often not required to view delayed feedback in previous classroom studies and they may have been less motivated to process the feedback when it was delivered after a delay. In Experiment 2, we manipulated whether or not feedback viewing was required in order to investigate whether this factor was responsible for producing an advantage of immediate feedback in previous classroom studies. Of additional interest was how often students would view the feedback if doing so was optional, and whether their viewing behavior would differ as a function of the timing of feedback.

4.1. Methods

Participants. Fifty students consented to participate in the experiment by releasing their data from the course for research purposes.

Design. The experiment used a 2 (timing of feedback: immediate, delayed) × 2 (feedback viewing: required, optional) mixed factorial design. Timing of feedback was manipulated within-subjects, but between-materials. The feedback-viewing requirement was manipulated between-subjects: students in one section were required to view feedback, while students in the other section were not.

Materials. The materials were the same as in Experiment 1 except that the instructor added one new topic (and corresponding homework assignment) for a total of nine topics and gave only three exams.

Procedure. The procedure was the same as Experiment 1 with a few exceptions. In each section of the class, students were split into two groups. Each group alternated from week to week between immediate and delayed feedback. Within each section, one group started with immediate feedback in the first week, whereas the other group started with delayed feedback. Thus, in any given week, some students in each section received immediate feedback while others received delayed feedback. As in Experiment 1, immediate feedback was released immediately after the assignment deadline,

[1] Corrected for violation of the equality of variances assumption.

whereas delayed feedback was released 1 week after the deadline. In addition, students in one section were required to view the feedback in order to get credit for the homework (as in Experiment 1), whereas students in the other section could choose whether or not to view feedback. Regardless of whether feedback viewing was required or optional, students received an email notification when the feedback was released.

4.2. Results

We applied the same criteria as in Experiment 1 to determine whether a student's data would be included in the analyses. Four students dropped the class before the first exam and thus they did not contribute any usable data. Of the remaining 46 students who completed at least one exam, 9 students were excluded from data analysis because they failed to complete at least half of all homework assignments (5 or more of 9) and/or exams (2 or more of 3). Most of the students who were excluded because of these criteria had dropped the course after the first exam. One additional student was excluded because the student registered in OpenStax Tutor for the wrong section of the course and thus did not receive the correct experimental conditions. After applying these criteria, a total of 36 students were included in the data analysis. We also excluded homework assignments (and their corresponding exam problems) according to same criteria as in Experiment 1. Applying these criteria resulted in the exclusion of the data related to 12% (38 out of 324) of homework assignments.

Homework performance. The proportion of correct multiple-choice responses on the homework assignments was approximately equal across the two sections (overall mean = 0.70), indicating that the two groups did not differ in knowledge prior to the implementation of the feedback manipulation. A 2 × 2 repeated measures ANOVA revealed no significant main effects of timing or feedback viewing (Fs < 1), and the interaction also was not significant [$F(1, 34) = 1.37$, MSE = 0.01, $p = 0.25$, $\eta^2 = 0.04$].

Feedback viewing. When feedback viewing was required, students viewed the feedback for 94% of all problems (immediate feedback = 93%; delayed feedback = 95%). In contrast, when feedback viewing was optional, students only viewed the feedback for 47% of all problems (immediate feedback = 53%; delayed feedback = 45%). A 2 × 2 repeated measures ANOVA revealed a significant main effect of feedback-viewing requirement [$F(1, 34) = 30.38$, MSE = 1186.78, $p < 0.01$, $\eta^2 = 0.47$]; neither the main effect of feedback timing ($F < 1$) nor the interaction were significant [$F(1, 34) = 2.52$, MSE = 194.03, $p = 0.12$ $\eta^2 = 0.07$]. When feedback viewing was required, students viewed the feedback an average of 5.8 days after completing the corresponding homework assignment in the immediate feedback condition and 14.3 days in the delayed feedback condition. When feedback viewing was optional, the corresponding averages were 13.0 days and 20.1 days, respectively.

Fig. 2 depicts the average number of times that students viewed the feedback for a given homework problem as a function of timing of feedback and feedback viewing requirement (excluding problems for which students did not view feedback at all). When students viewed the feedback, they did so a greater number of times if it was released immediately after the assignment deadline and they were required to view it. A 2 × 2 repeated measures ANOVA confirmed this observation by showing significant main effects of feedback timing [$F(1, 34) = 18.28$, MSE = 0.14, $p = 0.0002$, $\eta^2 = 0.35$] and feedback-viewing requirement [$F(1, 34) = 14.69$, MSE = 2.06, $p = 0.001$, $\eta^2 = 0.30$]; however, the interaction was not significant ($F < 1$).

Exam performance. Fig. 3 shows the proportion of correct multiple-choice responses on the exams as a function of timing of feedback and feedback-viewing requirement. Replicating the results of Experiment 1, students performed better on the exam when they had learned the material with delayed feedback. In addition, requiring students to view feedback also improved performance. A 2 × 2 repeated measures ANOVA revealed significant main effects of timing of feedback [$F(1, 34) = 4.79$, MSE = 0.05, $p = 0.036$, $\eta^2 = 0.12$] and feedback-viewing requirement [$F(1, 34) = 7.85$, MSE = 0.11, $p = 0.008$, $\eta^2 = 0.19$]. Although the interaction was not significant ($F < 1$), it is interesting to note that the size of the timing of feedback effect when feedback viewing was required ($d = 0.57$) was about twice that when feedback viewing was optional ($d = 0.23$).

Survey responses. Twenty-nine of the 36 students filled out the final survey. Collapsing across section, almost all of the students indicated that they "highly preferred" immediate feedback (90%) and no students preferred delayed feedback. In terms of their

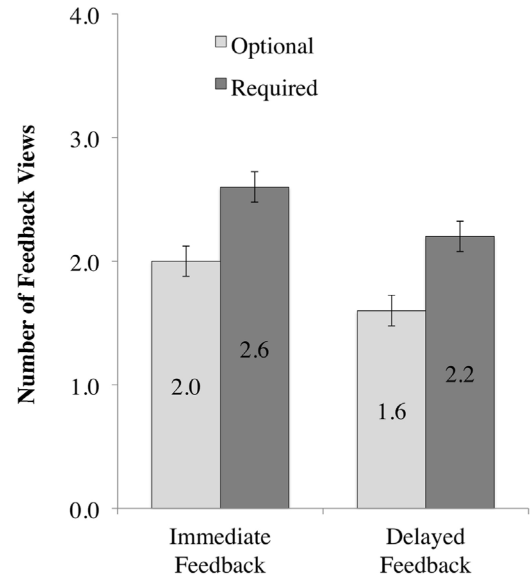

Fig. 2. Average number of times that a student viewed the feedback screen for a homework problem as a function of timing of feedback and feedback viewing condition given that the student viewed the feedback at least once. Error bars indicate 95% CI.

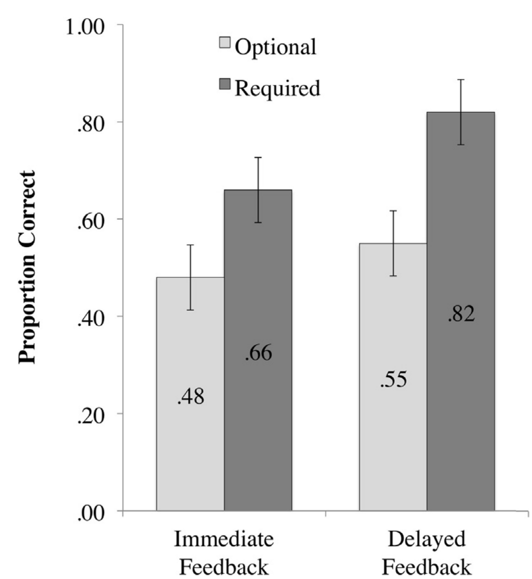

Fig. 3. Proportion of correct multiple-choice responses on the course exams as a function of timing of feedback and feedback viewing condition. Error bars indicate 95% CI.

performance in the course, most students reported that their performance "benefited a lot more" from immediate feedback (79%).

4.3. Discussion

Replicating the results of Experiment 1, delaying feedback had a positive effect on learning as measured by performance on the course exams. Requiring students to view feedback also improved learning, presumably because it motivated them to view the feedback; when feedback viewing was optional, students chose to view the feedback for only about half of the problems, which clearly hurt their performance on subsequent exams. Although a benefit of delaying feedback was evident even when feedback viewing was optional, the magnitude of the effect was much smaller. If students were more motivated to view the immediate feedback relative to delayed feedback in previous classroom studies, it is possible for this small effect to be eliminated or even reversed. In addition to replicating the effects of the timing of feedback on learning, we also found the same pattern of responses when we surveyed students in Experiment 2. Interestingly, despite having direct experience with both timing conditions, students still preferred immediate feedback and believed that it was more beneficial for their learning.

5. General discussion

In contrast to the common belief that feedback should be given immediately in classroom settings, we found that delaying feedback on the homework assignments enhanced the long-term retention and transfer of learning on exams in an undergraduate engineering course. This result also contrasted with the perceptions of students in the experiments who had direct experience with immediate and delayed feedback throughout the semester. Our findings add to a growing literature showing that delaying feedback can improve learning, and demonstrate that this effect is not limited to controlled laboratory settings. Indeed, it is impressive that this relatively small manipulation of feedback timing produced a measurable difference in student learning in this "noisy" classroom setting. Students engaged in a variety of other learning activities aside from the homework assignments, such as attending lectures, studying in groups, and meeting with teaching assistants. We made no attempt to control these extra-experimental factors, and yet the manipulation of feedback timing still had a significant effect on long-term retention and transfer.

Why did delaying feedback benefit learning? One possible explanation is that the viewing of delayed feedback took place closer in time to the exams than the viewing of immediate feedback. That is, the difference in exam performance may reflect a difference in retention interval rather than the timing of feedback per se (see Metcalfe et al., 2009). To investigate this possibility, we analyzed the data from the required viewing condition in Experiment 2 as a function of the retention interval from the assignment deadline to the exam. We classified the retention interval as "short" when the homework assignment was due 1 or 2 weeks before the corresponding exam and "long" when it was due 3 or 4 weeks beforehand. If the difference in retention interval explains our findings, one would expect to see the benefits of delayed feedback diminish over longer retention intervals. However, as Table 1 shows, the benefit of delayed feedback was approximately equal for both short and long retention intervals (if anything it was bigger at the long retention intervals). A 2 (feedback timing: immediate, delayed) × 2 (retention interval: short, long) ANOVA revealed a significant main effect of timing of feedback [$F(1, 15) = 5.56$, $MSE = 0.06$, $p = 0.03$, $\eta^2 = 0.12$]. Critically, however, neither the main effect of retention interval nor the interaction was significant ($F\text{s} < 1$). Although there was inadequate

Table 1
Average proportion of correct multiple-choice responses on the course exams in Experiments 1 and 2 as a function of short (1–2 weeks) versus long (3–4 weeks) retention interval between the homework assignment and corresponding exam. The means for Experiment 2 only represent performance in the required feedback viewing condition (the data from the optional feedback viewing condition were excluded from this analysis).

	Experiment 1		Experiment 2	
	Immediate	Delayed	Immediate	Delayed
Short (1–2 weeks)	.84	.92	.67	.81
Long (3–4 weeks)	.83	.94	.63	.81

power to conduct a similar analysis for Experiment 1, the data show the same pattern (Table 1). Thus, there is no evidence that a shorter retention interval in the delayed feedback condition accounted for the benefit of delayed feedback in the present study.

From a theoretical perspective, there are two possible explanations our findings, which are not mutually exclusive. The interference perseveration hypothesis holds that delaying feedback is beneficial because it allows learners to forget their errors, thereby reducing interference when learning the correct answers in the feedback message (Kulhavy & Anderson, 1972). Although this hypothesis has received some empirical support, there are also many findings that argue against it. For example, re-presenting learners with their initial errors as part of the delayed feedback message does not impair later memory for the correct information (e.g., Butler & Roediger, 2008). In addition, learners are often quite good at remembering their errors (e.g., Peeck & Tillema, 1979), and remembering an initial error can actually facilitate error-correction under some circumstances (Butler, Fazio, & Marsh, 2011). The interference perseveration hypothesis seems unlikely to explain the present results for at least one additional reason: the exams in this study consisted of transfer problems (i.e., different problems from those on the homework). Thus, it seems unlikely that errors from the homework problems would exert interference if they were remembered.

Another theoretical idea contends that the advantage of delaying feedback is a type of spacing effect, one of the oldest and most robust findings in the literature (for review see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Increasing the spacing between practice sessions typically results in better performance on a final retention test, with the optimal spacing interval depending on the length of the retention interval (i.e., how long the information has to be remembered). In the case of feedback, practice on the material is spaced or distributed over time when feedback is delayed, whereas it is massed all at once when feedback is given immediately. In situations where "immediate feedback" is not actually given directly after answering a question, such as in our study, a benefit of delaying feedback would be more appropriately described as a lag effect (see Delaney, Verkoeijen, & Spirgel, 2010). The lag effect refers to the finding that longer spacing intervals tend to produce better long-term retention (e.g., Glenberg, 1976). The spacing/lag effect explanation is most directly applicable to situations in which delaying feedback benefits the retention of initially correct responses. Given the relatively high level of performance on the initial homework problems in our experiments, it is likely that the students benefited from delayed feedback primarily because it constituted spaced practice at a greater interval. More generally, it is less clear how a spacing/lag effect explanation can account for the advantages of delaying feedback that sometimes emerge for initially incorrect responses (e.g., Butler & Roediger, 2008; Butler et al., 2007).

5.1. Limitations and future directions

Although the present study clearly demonstrates that delaying feedback does not impair student learning, the implications for the effectiveness of truly "immediate" feedback are less clear. In the present experiments, the immediate feedback was released to students after the assignment deadline rather than after each problem. The choice of this particular timing for immediate feedback was driven by the need to protect against student cheating, which is likely to be a concern in many classrooms. Nevertheless, future research should explore the relative effectiveness of feedback that is given immediately after the completion of a problem to feedback given at various delays. We suspect that there are situations in which receiving immediate feedback would be as beneficial or perhaps even more beneficial than receiving delayed feedback. For example, if students are going to be engaging in massed practice on a set of material, then delaying feedback might not make sense; it may be better to provide immediate feedback rather than allow the students to continue struggling with that task. Similarly, if students are less likely to view the feedback message after a delay (as we found in our research), then providing feedback directly after each item or after the test may produce superior learning. That said, there are many ways to ensure that students view feedback after a delay, such as reviewing feedback in a subsequent class or having students grade their own work.

The fact that the present study was conducted in an undergraduate classroom is a strength of this work, but these realistic conditions also resulted in a lack of control over various factors that might affect student performance. For example, students could seek out feedback on their own after completing the online homework assignments (e.g., talking to other students, consulting a textbook, etc.). We do not know how often this behavior occurred or whether it differed as a function of feedback-timing condition. Students might be more likely to seek out alternative sources of feedback when the feedback provided by the instructor is delayed. If so, the overall amount of feedback that a student receives (both self-initiated and instructor-provided) might be greater when feedback is delayed relative to when it is given immediately by the instructor. Future research should investigate the extent to which students make an effort to obtain feedback on their own, and how the timing of instructor-provided feedback affects this behavior.

5.2. *Practical applications*

Although there is clearly no single optimal time at which to provide students with feedback, we hope that our findings begin to change the common assumption that delaying feedback hinders learning. Our research shows that students share this misconception with educators and researchers. Student perceptions about how the timing of feedback affected their learning drastically departed from how it actually affected their performance on course exams. Prior research has shown that learners are often unable to discern which strategies are most effective for their learning even after direct experience (e.g., Bjork et al., 2013; Kornell & Son, 2009; Roediger & Karpicke, 2006). Our findings add to this substantial literature by demonstrating another metacognitive illusion that operates in the classroom.

With respect to educational practice, our findings speak to the importance of providing students with feedback and encouraging them to view it. Technology can be a useful tool for automatizing the delivery of feedback and tracking student feedback viewing. For example, the instructor in our study was able to get students to view feedback for almost every homework problem by using OpenStax Tutor and requiring students to view the feedback for course credit. Nevertheless, as the examples given above suggest, technology is not required to ensure that students view feedback after a delay. Regardless of how feedback is delivered to students, the message to educators is clear: when you take a few extra days to grade student work, feel free to tell them it is for their own good.

Conflicts of interest statement

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors would like to thank J.P. Slavinsky, Kim Davenport, Kevin Burleigh, and Andrew Ball for their contributions to this research. This research was supported by NSF Grant #IIS-1123617 to EJM. BV was supported by the Postdoctoral Fellowship for Teaching Excellence and Innovation at the University of Texas — El Paso. RvB was supported by NSF Grant #1041362.

All authors contributed to the idea for the research. HGM, ACB, and EJM designed the experiments. BV and RvB created the materials and taught the course. HGM and ACB analyzed the data. HGM drafted the manuscript. All authors edited the manuscript.

Appendix A.

Responses to survey questions in Experiments 1 and 2.

Experiment 1

Question	Timing of feedback	
	Immediate (N = 11)	Delayed (N = 7)
How did you feel about receiving feedback [immediately/after a delay]?		
I liked it a lot	82%	29%
I liked it a little	18%	14%
I did not like it	0%	57%
In terms of your performance in the course, do you feel like you benefited from [immediate/delayed] feedback?		
I benefited a lot	73%	29%
I benefited a little	27%	14%
I did not benefit	0%	43%
It hurt my performance	0%	14%
When you received feedback, how did you use it to help you learn?		
I did not look at it	0%	0%
I looked at how many problems I got correct/incorrect	0%	14%
I looked at the solution for problems that I got incorrect	64%	57%
I looked at the solution for all problems	36%	29%

Experiment 2

Question	Feedback Viewing	
	Required (N = 17)	Optional (N = 12)
Did you prefer receiving feedback immediately or after a delay?		
I highly preferred immediate feedback	94%	83%
I slightly preferred immediate feedback	0%	8%
I did not have a preference	6%	8%
I slightly preferred delayed feedback	0%	0%
I highly preferred delayed feedback	0%	0%
In terms of your performance in the course, do you feel like you benefited more from either immediate feedback or delayed feedback?		
I benefited a lot more from immediate feedback	76%	83%
I benefited a little more from immediate feedback	6%	0%
I benefited equally from immediate and delayed feedback	18%	8%
I benefited a little more delayed feedback	0%	8%
I benefited a lot more from delayed feedback	0%	0%
On weeks when you received immediate feedback, how did you use the feedback to help you learn?		

Question	Feedback Viewing	
	Required (*N* = 17)	Optional (*N* = 12)
I did not look at it	6%	0%
I looked at how many problems I got correct/incorrect	6%	8%
I looked at the solution for problems that I got incorrect	29%	50%
I looked at the solution for all problems	59%	42%
On weeks when you received delayed feedback, how did you use the feedback to help you learn?		
I did not look at it	18%	33%
I looked at how many problems I got correct/incorrect	12%	17%
I looked at the solution for problems that I got incorrect	24%	33%
I looked at the solution for all problems	47%	17%

References

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*, *63*, 186–188.

Angell, G. W. (1949). The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. *Journal of Educational Research*, *42*, 391–394.

Atkinson, R. C. (1969). Information delay in human learning. *Journal of Verbal Learning & Verbal Behavior*, *8*, 507–511.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, *13*, 111–128.

Battig, W. F., & Brackett, H. R. (1961). Comparison of anticipation and recall methods in paired-associate learning. *Psychological Reports*, *9*, 59–65.

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.

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

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

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

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.

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

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.

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*, *39*, 1211–1221.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 53) (pp. 63–147). New York: Academic Press.

Epstein Educational Enterprises. (2013). *What is the IF-AT?* Retrieved from http://www.epsteineducation.com/home/about/

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*, 1–16.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112.

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior: A program for self-instruction*. New York: McGraw-Hill.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Oxford, England: Appleton-Century.

Jensen, B. T. (1949). An independent-study laboratory using self-scoring tests. *Journal of Educational Research*, *43*, 134–137.

Kornell, N. (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*, 1297–1317.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*, 493–501.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*, 211–232.

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

Kulik, J. A., & Kulik, C. I. (1988). Timing of feedback and verbal learning. *Review of Educational Research Journal*, *21*, 79–97.

Landsman, H. J., & Turkewitz, M. (1962). Delay of knowledge of results and performance on a cognitive task. *Psychological Reports*, *11*, 66.

Little, J. K. (1934). Results of use of machines for testing and for drill upon learning in educational psychology. *Journal of Experimental Education*, *3*, 45–49.

Markowitz, N., & Renner, K. E. (1966). Feedback and the delay-retention effect. *Journal of Experimental Psychology*, *72*, 452–455.

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

Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum.

Paige, D. D. (1966). Learning while testing. *Journal of Educational Research*, *59*, 276–277.

(2013). *Pedagogical foundations*. Retrieved from https://www.coursera.org/about/pedagogy/

Pedagogy in action. (2013). *Classroom response systems*. Retrieved from http://serc.carleton.edu/sp/library/classresponse/index.html

Peeck, J., & Tillema, H. H. (1979). Learning from feedback: Comparison of two feedback procedures in a classroom setting. *Perceptual and Motor Skills*, *48*, 351–354.

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

Pressey, S. L. (1963). Teaching machine (and learning theory) crisis. *Journal of Applied Psychology*, *47*, 1–6.

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

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

Saltzman, I. J. (1951). Delay of reward and human verbal learning. *Journal of Experimental Psychology*, *41*, 437–439.

Sassenrath, J. M., & Yonge, G. D. (1968). Delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology*, *59*, 69–73.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Oxford, England: Appleton-Century.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24*, 86–97.

Skinner, B. F. (1958). Teaching machines. *Science*, *128*, 797–969.

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

Stephens, A. L. (1960). Certain special factors involved in the law of effect. In A. A. Lumsdaine, & R. Glaser (Eds.), *Teaching machines and programmed learning: A source book* (pp. 89–93). Washington, DC: National Education Association.

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

Sullivan, H. J., Schutz, R. E., & Baker, R. L. (1971). Effects of systematic variations in reinforcement contingencies on learner performance. *American Educational Research Journal*, *8*, 135–141.

Thorndike, E. L. (1932). *The fundamentals of learning*. New York, NY, US: Teachers College Bureau of Publications.

White, K. (1968). Delay of test information feedback and learning in a conventional classroom. *Psychology in the Schools*, *5*, 78–81.