American Psychological Association
Scholarship of Teaching and Learning in Psychology

2332-2101/19/$12.00

2019, Vol. 5, No. 4, 255–264
http://dx.doi.org/10.1037/stl0000140

Test Feedback and Learning: Student Preferences and Perceived Influence

Alyssa C. Smith, Brandon C. W. Ralph, Colin M. MacLeod, and Daniel Smilek
University of Waterloo

We examined student preferences for different forms of quiz feedback in the classroom and the perceived influence of these forms of feedback on learning. In 2 equivalent sections of an undergraduate university course in different semesters, students answered multiple-choice quiz questions about lecture content at the end of each class. The instructor provided 1 of 3 forms of feedback following each quiz question: (a) no feedback, (b) correct response feedback only, or (c) elaborative feedback consisting of the correct responses plus representation of the source material. Via online surveys both at the beginning and at the end of the semester, participants indicated their preferences for each type of feedback and their perceptions of the feedback's influence on their learning of the material. Elaborative feedback was preferred over feedback containing only the correct response, and both were preferred over receiving no feedback. Participants reported the same ordering with respect to their perceived learning of the lecture content and even reported that they found receiving no feedback harmful to their learning. Implications for student engagement and for testing effects are discussed.

Keywords: multiple choice, feedback, assessment, teaching/learning strategies

Supplemental materials: http://dx.doi.org/10.1037/stl0000140.supp

Multiple-choice testing, so frequently used in higher education (Bailey, Mossey, Moroso, Cloutier, & Love, 2012; DiBattista & Kurzawa, 2011), allows targeting of specific material as well as coverage of a wide range of topics from a course. Of course, the frequent use of multiple-choice tests is also partly because questions can be administered quickly and scored easily, virtues that are particularly valuable in large classes. With the advent of electronic response tools (e.g., smartphones and iClickers), multiple-choice testing can now also be implemented in most classroom settings, permitting more frequent and more immediate testing. For example, instructors can test students' knowledge of what was just taught by presenting them with multiple-choice questions on a slide at the end of a lecture; students then respond by clicking-in their answers using their electronic response tools. Immediately after their responses are collected, the instructor can provide feedback about the accuracy of response options. And, ordinarily, feedback is beneficial, particularly in the classroom (see Butler & Woodward, 2018; Hattie & Timperley, 2007; Kulik & Kulik, 1988).

Regularly using multiple-choice testing in classes might become an even more prevalent and important pedagogical strategy, given the growing body of work showing that testing while learning often improves memory for the material being learned (the testing effect; e.g., Butler & Roediger, 2008; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006; Roediger & Marsh, 2005). That is, practicing retrieval while

This article was published Online First March 28, 2019.
Alyssa C. Smith, Brandon C. W. Ralph, Colin M. MacLeod, and Daniel Smilek, Department of Psychology, University of Waterloo.
This research was supported by a Learning Innovation and Teaching Enhancement seed grant from the Centre for Teaching and Learning at the University of Waterloo awarded to Daniel Smilek and Colin M. MacLeod.
Correspondence concerning this article should be addressed to Alyssa C. Smith, Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada. E-mail: alyssa.smith@uwaterloo.ca

studying can lead to improved retention of the study material. Testing provides an opportunity for students to receive feedback about their performance and to discriminate the material that they know from the material that they do not know, allowing them to direct their studying toward the material not yet learned (Roediger et al., 2006). Moreover, the testing effect is quite robust, with the memory benefits of testing occurring in the laboratory when multiple episodes of testing are included (Roediger et al., 2006) or when a single episode of testing is used (Butler et al., 2008); the benefits also extend beyond the laboratory and into actual classroom settings (McDaniel, Anderson, Derbish, & Morrisette, 2007).

Testing while learning might also improve attentional focus on the material that is being learned. This was demonstrated by Szpunar, Khan, and Schacter (2013), who implemented an interpolated testing protocol in which participants watched a video lecture and then completed a memory test following the video. After each quarter of the lecture, some participants received a brief memory test, whereas others were either given an opportunity to restudy the material or were given mathematical problems to complete. In addition, during the lecture, participants responded to periodic thought probes asking them whether they had been focusing on the lecture or mind wandering just before the probe was presented. Strikingly, the results demonstrated that interpolated testing led to the least mind wandering and the best performance on the final test relative to the two other conditions in the study, suggesting that testing improves attentional focus on, and memory of, the lecture material.

In the real world, however, students might not be aware that testing can enhance their learning. Some indication that this is the case comes from a study reported by Karpicke et al. (2008), who had two groups of participants learn Swahili and English word pairs. One group practiced retrieval, whereas the other group did not. Interestingly, when asked to predict how much of the material they would recall a week later, both groups predicted the same amount (~50% of the word pairs). In actuality, the group that practiced retrieval recalled significantly more words a week later than the group that did not practice retrieval. One implication of this pattern of findings is that most participants seem to be unaware of the benefits of practicing retrieval while learning. This is consistent with Karpicke, Butler, and Roediger's (2009) survey indicating that only 11% of students practice retrieval while studying, with only 1% of students reporting that they practice retrieval as their primary study strategy.

Although there is considerable evidence that testing can improve learning despite students not being aware of it, the beneficial effects of such testing might depend on what happens after the test—specifically, whether students are given feedback regarding the accuracy of their responses. Indeed, one concern is that when no feedback is provided following a multiple-choice test, students may encode false knowledge (i.e., one of the incorrect answers) into memory (Butler et al., 2008; Roediger et al., 2005). This is because repeating information—whether that information is accurate or inaccurate—increases the likelihood that the repeated information will be endorsed as true when experienced subsequently (Bacon, 1979; Begg, Armour, & Kerr, 1985; Hasher, Goldstein, & Toppino, 1977). In the context of multiple-choice tests, when reading the various response options, students are essentially rehearsing both correct and incorrect answers (e.g., one correct answer and three incorrect answers in the case of a four-alternative question). Consequently, when students receive no feedback regarding the correct response, they may inadvertently encode incorrect information. Later, when source memory has eroded over time, this incorrect information might be treated as being accurate. Fortunately, providing students with feedback regarding the correct response in such testing scenarios counteracts the negative effects of encoding false lures during multiple-choice tests (Butler et al., 2008). For example, Butler, Karpicke, and Roediger (2007) had participants study passages and then complete two multiple-choice tests—one immediately after studying and one the following day. Participants received no feedback, feedback immediately following each question, or delayed feedback (10 min later or the next day). Receiving feedback, regardless of delay, resulted in significantly better memory performance the next day compared with not receiving feedback.

Although the benefits of feedback (and detriments of no feedback) have been shown objectively in terms of memory performance, student attitudes toward, and perceptions of the utility of, such feedback remain to be explored. We posit that such attitudes toward, and perceptions of, test feedback are important because they likely drive broad behaviors, such as the decision to seek out feedback and to effectively engage with it. For example, students may be unlikely to seek out feedback if they do not interpret such feedback as beneficial. In fact, there is evidence that in some cases students do not request feedback (Schloss, Sindelar, Cartwright, & Smith, 1988) and that they do not review all of the feedback that they receive (Mullet, Butler, Verdin, von Borries, & Marsh, 2014). In addition, even when students are presented with feedback such that seeking out feedback is not an issue, their attitudes about feedback might prevent them from engaging with the feedback in a useful or mindful way (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Shute (2008), for instance, has speculated that students may fail to engage with feedback when they interpret it as being too simple, too complex, or too nonspecific. Accordingly, we suggest that it is important to examine students' attitudes about and perceptions of the utility of feedback.

The Present Study

Given the foregoing considerations, in the present study, we sought to investigate students' views of test feedback, both in terms of their preferences for receiving different forms of feedback and their beliefs about the utility of different forms of feedback. Here the forms of feedback evaluated included: (a) no feedback, (b) correct response–only feedback, and (c) elaborative feedback, consisting of the correct response plus a representation of the lecture slide pertinent to that question's content. Across two samples (i.e., two academic semesters), students responded to multiple-choice test (quiz) questions at the end of most lectures in an undergraduate university course.[1] On each day and following each quiz question, participants received one of the three forms of feedback. Critically, through online surveys administered both at the beginning and at the end of each semester, students indicated (a) their preference for each form of feedback and (b) their perception of the influence that each form of feedback had on their learning.

Method

Following the recommendations of Simmons, Nelson, and Simonsohn (2012), we report how we determined our sample size, all manipulations, all measures, and all data exclusions in this study.

Participants

All participants were enrolled in a second-year (Physiological Psychology) course, either during a Fall (Sample 1) or Winter (Sample 2) academic term and received partial course credit in exchange for participation. Given that all students enrolled in the course were eligible to participate, we did not have an a priori targeted sample size. Instead, we collected data from as many participants as agreed to take part in the study (limited only by the enrollment capacity of the course). Because we surveyed students' preference ratings and perceived influence ratings at the beginning and at the end of the term, there were some participants who completed the beginning of term questionnaire but did not complete the end of term questionnaire. In Sample 1, 143 students completed the beginning of term questionnaire, and 110 of these completed the end of the term questionnaire. In Sample 2, 95 students completed the beginning of term questionnaire, and 56 of these completed the end of term questionnaire. Only those students who completed both questionnaires are included in the data analyses.

To encourage participation in the study, at the beginning of the term, a 5-min presentation was given by one of the coauthors (BR), informing students of the details of the study. This presentation made clear to students that only data from those who consented to participate would be included in the study and that the instructor and teaching assistants would not have knowledge of who was or was not participating in the study until after grades were finalized.

Materials and Procedure

The classes. This study was conducted across two sections of the same undergraduate course (Physiological Psychology) at the University of Waterloo (taught by coauthor, DS).

[1] Quiz questions at the end of lecture contributed to a course participation grade.

At the end of each class meeting, as part of normal classroom activities, participants completed three to five quiz questions (depending on the time remaining at the end of each class) based on that day's lecture material. Quiz questions were displayed via a web-browser applet using an overhead projector to present the questions on a screen at the front of the class. Participants indicated their responses using electronic remotes (iClickers; www.iclicker.com) across 18 class meetings in Sample 1 and across 20 class meetings in Sample 2. In total, students responded to 74 quiz questions in Sample 1 and 66 quiz questions in Sample 2.

Form of feedback manipulation. In the focal aspect of the study, we manipulated the form of feedback that participants received following each quiz question. In the No Feedback condition, participants received no information following their responses. In the Correct Feedback condition, participants were told what the correct answer was after everyone in the class responded. And in the Elaborative Feedback condition, after responding to a quiz question, participants were provided the correct answer and were reshown the lecture slide containing the answer. These represented lecture slides were not altered in any way but were accompanied by a 20- to 60-s verbal comment. The form of feedback was pseudorandomly varied across lectures, with the constraint that the various types of feedback were approximately equally distributed across classes. In Sample 1, 27 No Feedback, 21 Correct Feedback, and 26 Elaborative Feedback quiz questions were presented. In Sample 2, 23 No Feedback, 20 Correct Feedback, and 23 Elaborative Feedback quiz questions were presented. The variation in question counts per feedback condition occurred because some questions were not presented as a result of time constraints in a given lecture.

Time-allotted manipulation. In addition to our primary manipulation of Level of Feedback, for exploratory purposes, we also manipulated the amount of time that participants had to respond to each question. For Fixed Time quiz questions, participants received 45 s to respond to the quiz question. Variable Time quiz questions varied between 30, 45, and 60 s based on the anticipated difficulty of the question (easy, medium, or hard, respectively).[2] All quiz questions within a given lecture fell under either the Fixed Time or Variable Time condition, and the timing conditions were approximately balanced across lectures. A timer showing the remaining time was presented on the screen while participants were answering the questions. In Sample 1, 34 Fixed Time and 40 Variable Time quiz questions were presented. In Sample 2, 34 Fixed Time and 32 Variable Time quiz questions were presented. Our motivation for including this manipulation of Time Allotted was to explore whether students preferred and believed there was a benefit to modularly timed test questions. Given that this manipulation was included for exploratory purposes and is not directly relevant to the purpose of the present article, data pertaining to this manipulation are included in the online supplementary material (see Appendix A).

Beginning and end-of-term questionnaire. Full copies of our beginning and end-of-term questionnaires appear in the online supplementary materials (see Appendix B). Preferences for quiz questions with each Form of Feedback, as well as metacognitions about their influence on learning, were assessed using self-reports. These self-reports were collected at the beginning of the term (within the first month), when students had relatively little experience with each of our manipulations, and at the end of the term, when students had considerably more experience with each of our manipulations.

Preferences ratings. Participants indicated their preference for each form of feedback by responding to the statement, "Please rate the extent to which you prefer end-of-class quizzes with each of the following testing styles": using a 6-point scale with anchors (1) *highly disliked*, (2) *moderately disliked*, (3) *slightly disliked*, (4) *slightly preferred*, (5) *moderately preferred*, and (6) *highly preferred*.

Perceived influence ratings. Similarly, participants indicated their perceptions of the effect of each testing style on their learning by responding to the statement, "Please rate the extent to which you feel end-of-class quizzes with each of the different testing styles helped you learn the course material" using a 6-point scale with anchors (1) *hindered my learning a lot*, (2) *moderately hindered*, (3) *slightly hin-*

[2] The form of feedback did not vary with question difficulty.

dered, (4) *slightly helped*, (5) *moderately helped*, and (6) *helped my learning a lot*.

Other measures. We also collected preference and perceived influence judgments for the Time Allotted (Fixed Time vs. Variable Time) manipulation, using the same statements and response options as outlined above. In addition, the questionnaire included items inquiring about the participant's age, gender, prior knowledge of the brain, motivation to learn, how frequently they completed assigned readings prior to lecture, and how long they typically needed to comfortably complete a multiple-choice quiz question. In Sample 2, we also included an exploratory mind-wandering questionnaire. These items are provided in full in the online supplementary materials (see Appendix B).

Results

Preference Ratings

Mean preference ratings, averaged across participants, are shown in Figure 1A (Sample 1) and Figure 1B (Sample 2). Preference ratings for each form of quiz feedback were evaluated using a 3 × 2 × 2 mixed analysis of variance, with Form of Feedback (No Feedback vs. Correct Feedback vs. Elaborative Feedback) and Time of Term (Start vs. End) entered as the within-subject variables, and Sample (Sample 1

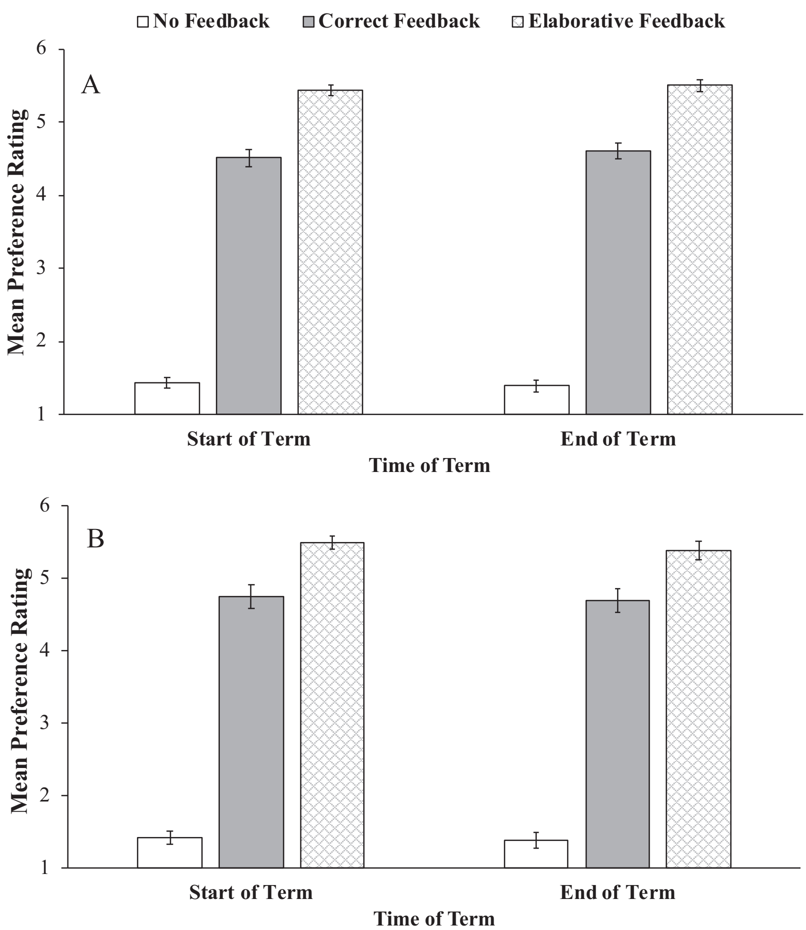

Figure 1. Mean preference ratings from Sample 1 (Fall; A) and Sample 2 (Winter; B) for each level of feedback. Error bars represent 1 SE of the corresponding mean.

vs. Sample 2) entered as the between-subjects variable. The analysis revealed a significant main effect of Form of Feedback, $F(2, 328) = 944.53$, $p < .001$, $\eta_p^2 = .86$. No other main effects or interactions reached significance, all Fs < 1.40, all ps $> .240$.

To further examine the main effect of Form of Feedback, we conducted three post hoc pairwise t tests (collapsing across Time of Term and Sample) comparing each level of Form of Feedback with each of the other two levels. These tests revealed that Elaborative Feedback was more preferred compared with both Correct Feedback, $t(165) = 8.51$, $p < .001$, and No Feedback, $t(165) = 53.37$, $p < .001$, and that Correct Feedback was more preferred than No Feedback, $t(165) = 31.58$, $p < .001$. Using Bonferroni corrections for multiple comparisons (p of 0.05 divided by three comparisons), these comparisons remained significant at a corrected significance criterion of $p = .017$.

Perceived Influence on Learning

The mean Perceived Influence on Learning ratings, averaged across participants, are shown in Figure 2A (Sample 1) and Figure 2B (Sample 2). Perceived Influence ratings were submitted to the same 3 × 2 × 2 mixed analysis of variance, with Form of Feedback (No Feedback vs. Correct Feedback vs. Elaborative Feedback) and Time of Term (Start vs. End) entered as the within-subject variables and Sample (Sample 1 vs. Sample 2) entered as the between-subjects variable. This analysis also revealed a significant main effect of Form of Feedback, $F(2, 326$[3]$) = 647.65$, $p < .001$, $\eta_p^2 = .80$, and no other main effects or interactions reached significance, all Fs < 1.00, all ps $> .319$.

Given that we again found only a main effect of Form of Feedback for perceived influence ratings and no interactions by sample, we collapsed across Time of Term and Sample for post hoc pairwise t tests. These tests revealed that Elaborative Feedback was perceived as more helpful than both Correct Feedback, $t(165) = 12.28$, $p < .001$, and No Feedback, $t(165) = 34.48$, $p < .001$. Correct Feedback also was perceived as more helpful than No Feedback, $t(165) = 24.01$, $p < .001$. Using Bonferroni corrections for multiple comparisons, these comparisons remained significant at the $p = .017$ level.

Pearson Correlations of Preference Ratings and Perceived Influence Ratings

We were also interested in whether there was a relation between preference and the perceived influence on learning ratings for each form of feedback. Because there was no main effect of Sample on feedback preferences or perceived influence ratings and no main effect of Time of Term, we collapsed across samples to increase sample size for these correlations. There were significant positive correlations between preference and perceiving influence ratings for No Feedback, $r(164) = .517$, $p < .001$, Correct Feedback, $r(164) = .682$, $p < .001$, and Elaborative Feedback, $r(163$[4]$) = .685$, $p < .001$. In other words, the extent to which students preferred any given feedback form was related to the extent to which they felt that it positively influenced their learning.

Discussion

In the current article, we sought to evaluate student preferences for, and their views about the utility of, different forms of feedback following multiple-choice testing in the classroom. Across two academic semesters, we found that students preferred to receive either correct-only or elaborative feedback, compared with no feedback and reported their belief that such feedback (correct-only and elaborative) positively influenced their learning of course material. Moreover, elaborative feedback was the most preferred and rated as having the greatest positive influence on learning. By contrast, ratings of the no feedback condition indicated that, on average, students disliked receiving no feedback and felt that receiving no feedback actually hindered their learning to some degree. Finally, it is worth noting that in both samples preference and perceived influence ratings remained stable over the course of the academic semester, indicating that these views either required little exposure or were already held by students before taking part in this study.

[3] One participant did not rate the Perceived Influence of Elaborative Feedback and as a result was removed from analyses including that variable.

[4] One participant did not rate the perceived influence of Elaborative Feedback; thus, this correlation includes 165 participants.

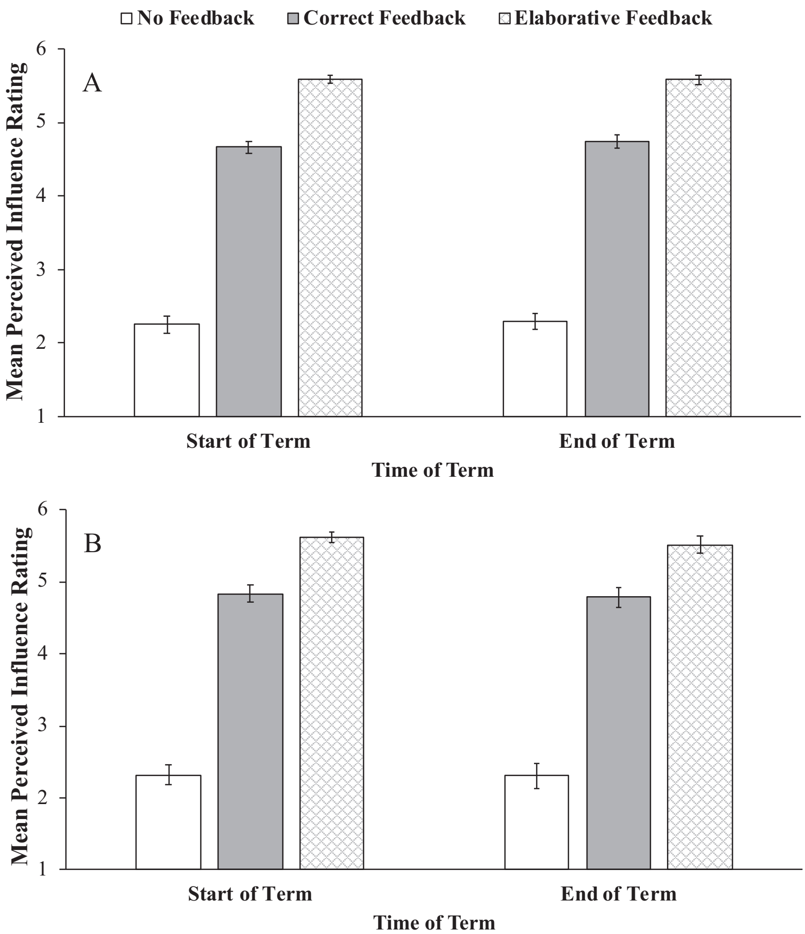

Figure 2. Mean perceived influence ratings from Sample 1 (Fall; A) and Sample 2 (Winter; B) for each form of feedback. Error bars represent 1 SE of the corresponding mean.

Our findings that students prefer and perceive a positive benefit for both correct and elaborative feedback are consistent with prior work demonstrating objective benefits of such feedback on learning (Butler, Godbole, & Marsh, 2013). Importantly, our findings suggest that students are explicitly aware that test feedback has a beneficial impact on learning. Moreover, we also found that students preferred, and reported a greater benefit from, elaborative feedback over correct response–only feedback. This subjective perception of elaborative over correct-only feedback is consistent with some findings in the literature demonstrating that elaborative feedback confers memory performance benefits over and above correct response–only feedback. For example, previous work on feedback and memory retention has shown that elaborative feedback improves memory over and above correct-only feedback on new inference questions, although this does not appear to be the case for retest performance of studied material (Butler et al., 2018; Butler et al., 2013). As suggested by our correlations, students might prefer elaborative feedback because of its learning benefits. However, another possibility is that students' perceptions of the benefits of feedback on learning may be driven by yet-unmeasured factors.

Another noteworthy observation is that, on average, students not only reported disliking receiving no feedback but also reported that receiving no feedback following quizzes hindered their learning of course material. Recall that mean influence ratings for no feedback were 2.22 and 2.21 in Samples 1 and 2, respectively, which fell between our *moderately hindered* (2) and *slightly hindered* (3) scale anchors. These subjectively reported detrimental effects are consistent with prior work demonstrating objective detrimental effects of multiple-choice testing when no feedback is given, whereby incorrect answers/lures become more likely to be encoded into memory (Butler et al., 2008; Roediger et al., 2005). Coupling the student-reported evaluations observed here with prior concerns raised in the literature, it appears prudent to provide timely, informative feedback following multiple-choice testing to avoid incurring costs to learning. Of course, always providing elaborative feedback can be time consuming (especially in a live classroom setting). Thus, one recommendation is that instructors minimally provide correct-only feedback and add elaborative feedback when possible. For instance, a reviewer of this article suggested that instructors might choose to provide elaborative feedback on only the most problematic questions from a test in class.

Given that the correlations between preference and influence ratings were relatively high (ranging between .52 and .69) yet not high enough to suggest that they are redundant, it is important to consider whether and how preferences and perceived utility might influence each other. In our study, we found that preference and perceived-influence ratings were positively correlated. On the one hand, it may be that subjective preferences for a particular form of feedback influence the extent to which individuals endorse that form of feedback as helpful or harmful with respect to learning. For example, Apperson, Laws, and Scepansky (2006) found that students perceived classes in which PowerPoint (vs. talk and chalk) was used as having a more positive influence on learning—such that students rated the PowerPoint classes as better structured, clearer, and more stimulating. There was no difference in grades as a result of the use of PowerPoint; however, students in the PowerPoint group rated a higher likelihood of taking another course with the same instructor. This suggests that while perceptions of utility do not necessarily predict objective learning outcomes, they may predict broader behavior. On the other hand, it may be the case that preferences for feedback are governed by the perceived influence of the feedback on learning. It makes intuitive sense, after all, for students to prefer a style of feedback that they feel benefits them most in terms of their learning goals. And obviously there is always the possibility that the strong relation between preferences and perceived utility scores in our study was caused by a common third variable. Because in the present work we did not set out to dissociate these possible causal relations, the nature of the causal relations remains an empirical question for future investigation.

Whereas our findings show that students like to receive feedback and find it useful, anecdotal experience and experimental evidence suggest that when students have to actively seek out feedback (e.g., attending the office hours of an instructor or teaching assistant to receive feedback on a midterm or final exam), many students do not exert that effort. For instance, Mullet et al. (2014) found that when students were not required to review the feedback that they had received on their homework assignments, they reviewed the feedback only about half the time. Students only reviewed the feedback consistently (about 90% of the time) when it was made a course requirement. Why is there this disconnect between students' views and their practice? One possibility is that, despite apparently understanding that feedback is beneficial, students are willing to seek out and engage with feedback only when the perceived benefits of doing so outweigh the costs (e.g., time commitment). Given these tradeoffs, perhaps the best solution is to actively incorporate feedback in the classroom because this way students do not have to exert extra effort to seek it out and because, as shown here, the classroom is a context in which students report preferring feedback (over no feedback) and perceive it to be helpful.

We conclude by noting that currently many standard forms of testing in the classroom—including midterms and exams—reflect the no feedback style of testing. This is understandable for a variety of reasons: Instructors do not always have time to provide detailed feedback following in-class quizzes, midterms, or final exams, and sharing the answers of assessments may not be desirable if tests rely on the same or similar question banks across repeated offerings. Nevertheless, we suggest that instructors ought to take opportunities to provide feedback when these opportunities arise. Building on prior studies showing that providing feedback improves learning, here we have shown that students prefer feedback over no feedback and they believe that feedback aids their learning.

References

Apperson, J. M., Laws, E. L., & Scepansky, J. A. (2006). The impact of presentation graphics on students' experience in the classroom. *Computers & Education, 47,* 116–126. http://dx.doi.org/10.1016/j.compedu.2004.09.003

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 241–252. http://dx.doi.org/10.1037/0278-7393.5.3.241

Bailey, P. H., Mossey, S., Moroso, S., Cloutier, J. D., & Love, A. (2012). Implications of multiple-choice testing in nursing education. *Nurse Education Today, 32,* e40–e44. http://dx.doi.org/10.1016/j.nedt.2011.09.011

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238. http://dx.doi.org/10.3102/00346543061002213

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioural Science, 17,* 199–214. http://dx.doi.org/10.1037/h0080140

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105,* 290–298. http://dx.doi.org/10.1037/a0031026

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. http://dx.doi.org/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. http://dx.doi.org/10.3758/MC.36.3.604

Butler, A. C., & Woodward, N. R. (2018). Towards consilience in the use of task-level feedback to promote learning. In K. D. Federmeier (Ed.), *The psychology of learning and motivation* (Vol. 69, pp. 1–38). Amsterdam, Netherlands: Elsevier.

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning, 2,* 1–23. http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16,* 107–112. http://dx.doi.org/10.1016/S0022-5371(77)80012-1

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. http://dx.doi.org/10.3102/003465430298487

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17,* 471–479. http://dx.doi.org/10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968. http://dx.doi.org/10.1126/science.1152408

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58,* 79–97. http://dx.doi.org/10.3102/00346543058001079

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19,* 494–513. http://dx.doi.org/10.1080/09541440701326154

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 & Cognition, 3,* 222–229. http://dx.doi.org/10.1016/j.jarmac.2014.05.001

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. http://dx.doi.org/10.1037/0278-7393.31.1.3

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1155–1159. http://dx.doi.org/10.1037/0278-7393.31.5.1155

Schloss, P. J., Sindelar, P. T., Cartwright, G. P., & Smith, M. A. (1988). Learner control over feedback as a variable in computer assisted instruction. *Journal of Research on Computing in Education, 20,* 310–320. http://dx.doi.org/10.1080/08886504.1988.10781845

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153–189. http://dx.doi.org/10.3102/0034654307313795

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue, 26,* 4–7.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 6313–6317. http://dx.doi.org/10.1073/pnas.1221764110

Received September 24, 2018
Revision received December 18, 2018
Accepted January 23, 2019 ■