

Explanation Feedback Is Better Than Correct Answer Feedback for Promoting Transfer of Learning

Andrew C. Butler
Duke University

Namrata Godbole
University of North Carolina–Greensboro

Elizabeth J. Marsh
Duke University

Among the many factors that influence the efficacy of feedback on learning, the information contained in the feedback message is arguably the most important. One common assumption is that there is a benefit to increasing the complexity of the feedback message beyond providing the correct answer. Surprisingly, studies that have manipulated the content of the feedback message in order to isolate the unique effect of greater complexity have failed to support this assumption. However, the final test in most of these studies consisted of a repetition of the same questions from the initial test. The present research investigated whether feedback that provides an explanation of the correct answer promotes superior transfer of learning to new questions. In 2 experiments, subjects studied prose passages and then took an initial short-answer test on concepts from the text. After each question, they received correct answer feedback, explanation feedback, or no feedback (Experiment 1 only). Two days later, subjects returned for a final test that consisted of both repeated questions and new inference questions. The results showed that correct answer feedback and explanation feedback led to equivalent performance on the repeated questions, but explanation feedback produced superior performance on the new inference questions.

Keywords: feedback, learning, retention, transfer

Feedback is a critical component of any learning process because it allows learners to reduce the discrepancy between actual and desired knowledge (Black & Wiliam, 1998). Although prior research has identified many factors that influence the efficacy of feedback (for reviews, see D. L. Butler & Winne, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), the content of the feedback message is arguably the most important aspect of any feedback procedure. The information supplied in the feedback message is critical because it enables learners to correct errors (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005) and maintain correct responses (e.g., A. C. Butler, Karpicke, & Roediger, 2008). Thus, a primary objective of feedback research is to determine what information the feedback message should contain in order to be maximally effective.

On the basis of a wealth of studies in the literature, the current answer is that the feedback message should contain the correct answer. At the most basic level, feedback must convey information about the veracity of the learner's response (i.e., correct vs. incorrect). However, many studies have shown little or no benefit of providing verification feedback relative to no feedback (e.g., Pashler et al., 2005; Plowman & Stroud, 1942; Roper, 1977; but see Fazio, Huelser, Johnson, & Marsh, 2010). Including the correct answer in the feedback message substantially increases the efficacy of feedback because it provides the information that learners need to correct their errors. Indeed, the vast majority of studies that have compared correct answer feedback with verification feedback have shown a superiority of correct answer feedback (e.g., Pashler et al., 2005; Phye & Sanders, 1994; Roper, 1977; Travers, Van Wagenen, Haygood, & McCormick, 1964; Whyte, Karolick, Neilsen, Elder, & Hawley, 1995).

Is it beneficial for the feedback message to include other information in addition to the correct answer? A common assumption among educators and researchers is that providing students with additional information in the feedback message will improve learning. The umbrella term *elaborative feedback* is often used to describe any type of feedback that is more complex than correct answer feedback, and there are many ways of elaborating the feedback message (for a taxonomy of feedback messages, see Kulhavy & Stock, 1989). Examples of elaborative feedback include providing an explanation of why a particular response is correct or incorrect (*explanation feedback*) and re-presenting the original learning materials (*restudy feedback*). Due to the assump-

This article was published Online First December 17, 2012.

Andrew C. Butler, Psychology and Neuroscience, Duke University; Namrata Godbole, Department of Psychology, University of North Carolina–Greensboro; Elizabeth J. Marsh, Psychology and Neuroscience, Duke University.

This research was supported by a Collaborative Activity Award from the James S. McDonnell Foundation's 21st Century Science Initiative in Bridging Brain, Mind and Behavior awarded to Elizabeth J. Marsh. We thank Katherine Rawson and the members of the Marsh Lab for their helpful comments and suggestions on drafts of the article.

Correspondence concerning this article should be addressed to Andrew C. Butler, Psychology and Neuroscience, Duke University, Box 90086, Durham, NC 27708-0086. E-mail: andrew.butler@duke.edu

tion that elaborative feedback is helpful to students, it is often included as a component in methods of instruction, such as intelligent tutoring systems (Corbett, Koedinger, & Anderson, 1997) and computer-assisted instruction programs (Gibbons & Fairweather, 1998). For instance, the AutoTutor is an intelligent tutoring system that helps learners solve complex physics problems by providing many different types of feedback, including hints, corrections, and explanations (Graesser, Chipman, Haynes, & O'Leary, 2005). However, elaborative feedback is just one of many components combined to enhance student learning in such systems, and its independent contribution to learning is not assessed.

Surprisingly, studies that have directly compared elaborative feedback with correct answer feedback have found little or no benefit to increasing the complexity of the feedback message (for a review, see Kulhavy & Stock, 1989; for a meta-analysis, see Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). For example, many studies have found that there is no benefit of providing explanation feedback relative to correct answer feedback (e.g., Gilman, 1969; Kulhavy, White, Topp, Chan, & Adams, 1985; Mandernach, 2005; Pridemore & Klein, 1995; Sassenrath & Gaverick, 1965; Smits, Boon, Sluijsmans, & van Gog, 2008; Whyte et al., 1995). Similarly, other studies have shown that providing restudy feedback yields equivalent performance to correct answer feedback (e.g., Andre & Thieman, 1988; Kulhavy et al., 1985; Peeck, 1979). Critically, the content of the feedback message was manipulated as an independent variable in these studies, which allowed the unique effect of greater complexity (or lack thereof) to be isolated.

The lack of empirical support for the efficacy of elaborating the feedback message is surprising, but these null effects may be due to how learning was assessed on the final test. Almost all of the studies on elaborative feedback have used a final test that assessed retention of the correct answer by repeating the same questions from the initial test. If the learner only needs to remember the correct answer to perform well on the final test, then the additional information contained in elaborative feedback is superfluous. However, this additional information may be important for fostering better comprehension of the material. For example, providing an explanation of why a response is correct (i.e., explanation feedback) might help the learner to move from superficial factual knowledge to a more complex understanding of the concept. Thus, elaborative feedback might be expected to facilitate performance on a final test that assesses understanding rather than retention of the correct response. One hallmark of superior understanding is the ability to transfer knowledge to new contexts. *Transfer* can be broadly defined as "the influence of prior learning (retained until the present) upon the learning of, or response to, new material . . ." (McGeoch, 1942, p. 394). In the present study, we assess understanding by investigating learners' ability to transfer their knowledge on a final test that involves making inferences using previously learned concepts.

Experiment 1

The goal of the first experiment was to investigate the hypothesis that the efficacy of elaborative feedback depends on how learning is assessed. Subjects studied a set of passages and then took an initial test on critical concepts from the passages. After each question, they received explanation feedback, cor-

rect answer feedback, or no feedback. Two days later, they took a final test that assessed both retention (via repeated questions from the initial test) and transfer (via new inference questions). We predicted that explanation feedback would lead to better transfer relative to correct answer feedback, but the two types of feedback would produce equivalent retention of the correct answers.

Method

Participants. Sixty Duke University students participated for either course credit or payment. Four additional subjects were excluded because they failed to follow experimental instructions.

Design. The experiment had a 3 (type of feedback: no feedback, correct answer, explanation) \times 2 (type of final test question: repeated, new) mixed factorial design. Type of feedback was manipulated between subjects, and type of final test question was manipulated within subjects, between materials.

Materials and counterbalancing. Materials consisted of 10 passages about a variety of topics (e.g., the respiratory system, tropical cyclones, etc.) and associated questions. Six of the passages and the associated questions were adapted from A. C. Butler (2010), and the rest were created to match. Each passage consisted of 500 words of text and contained two critical concepts (see Appendix A for sample passages). Thus, there was a total of 20 critical concepts. A *concept* was operationally defined as a piece of information that must be abstracted from multiple sentences. Two questions were associated with each concept: a definition question and an inference question. All of the definition questions were used on the initial test to assess memory for the 20 concepts. The definition question was repeated on the final test for 10 of the concepts in order to assess retention of the correct answer. The inference question was given on the final test for the other 10 concepts in order to assess transfer of knowledge. The materials were counterbalanced by creating two versions of the final test. In each version, one of the two concepts for each passage was tested by repeating the definition question, whereas the other concept was tested with a new inference question. Thus, each concept was tested equally often in each final test condition across subjects.

Two feedback messages were created for each definition question: a correct answer message and an explanation message. The correct answer message consisted of a statement of the correct answer, whereas the explanation message consisted of the correct answer as well as two additional sentences elaborating on the correct answer. The two additional sentences in the explanation feedback message were taken from the passage and helped to explain the concept. The explanation feedback did not contain any new information and it did not provide the answer to the inference question. Appendix B contains sample questions and feedback.

Procedure. The experiment consisted of two sessions spaced 2 days apart. Individual PCs running MediaLab software (Jarvis, 2004a, 2004b) were used to present all the materials and collect the responses. In Session 1, subjects were randomly assigned to one of the three feedback conditions (no feedback, correct answer feedback, and explanation feedback). Regardless of condition, they studied the 10 passages in a random order determined by the computer. Each passage was divided into two paragraphs, and each paragraph was presented for 80 s (pilot testing showed this amount

of time to be sufficient to read the entire paragraph once). Next, subjects engaged in a distractor task for 5 min (solving visuospatial puzzles). Finally, they completed a self-paced short-answer test on the critical concepts that consisted of the 20 definition questions. The questions were presented one at a time in a random order, and subjects were required to generate a response to each question. If they did not know the answer, they were instructed to make a plausible guess. Immediately after answering each question, subjects received the type of feedback that they had been assigned (no feedback, correct answer feedback, or explanation feedback). The question was always re-presented with the feedback message to provide context. Feedback was provided regardless of whether the response was correct or incorrect, and subjects were required to study the message for 20 s. In Session 2, subjects returned after 2 days to take a final short-answer test that contained 20 questions: 10 definition questions that were repeated from the initial test and 10 new inference questions. As on the initial test, questions were presented one at a time in a random order, answering was self-paced, and subjects were required to respond to each question.

Results

All results were significant at the .05 level unless otherwise stated. Pairwise comparisons were Bonferroni-corrected to the .05 level. Eta-square and Cohen's d are the measures of effect size reported for all significant effects in the analysis of variance (ANOVA) and t -test analyses, respectively.

Scoring. Two coders independently coded the responses as correct or incorrect according to a scoring rubric. Both coders were blind to condition and coded all the responses for a given question together to increase consistency. Cohen's kappa was calculated to assess interrater reliability. Reliability was high ($\kappa = .89$), and the first author (ACB) resolved the disagreements in scoring.

Initial test performance. Initial test performance was relatively low (grand $M = .43$), which was desirable for investigating the effects of feedback. A one-way ANOVA showed no effect of feedback condition ($F < 1$).

Final test performance. Figure 1 shows the proportion of correct responses on the final test as a function of feedback condition on the initial test and type of final test question. When subjects received correct answer or explanation feedback on the initial test, they performed better on the repeated definition questions relative to when they did not receive feedback. A one-way ANOVA confirmed this observation by revealing a significant main effect of type of feedback, $F(2, 57) = 6.54$, $MSE = .05$, $\eta^2 = .19$. Follow-up pairwise comparisons showed that both the correct answer and explanation feedback conditions led to a greater proportion of correct responses on repeated questions relative to the no-feedback condition (.62 vs. .43), $t(38) = 2.63$, $SED = .07$, $d = .85$; and (.66 vs. .43), $t(38) = 3.34$, $SED = .07$, $d = 1.06$, respectively. However, there was no significant difference between the correct answer and explanation feedback conditions ($t < 1$).

On the new inference questions, subjects performed best when they had received explanation feedback relative to when they got correct answer or no feedback. A one-way ANOVA showed a significant main effect of type of feedback, $F(2, 57) = 6.55$, $MSE = .04$, $\eta^2 = .19$. Pairwise comparisons confirmed that the explanation feedback condition produced a significantly greater

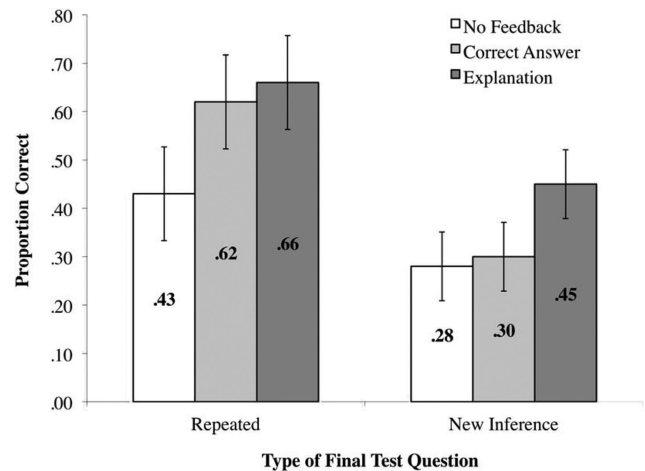


Figure 1. Proportion of correct responses on the final test as a function of feedback condition on the initial test and type of final test question in Experiment 1. Error bars represent 95% confidence intervals.

proportion of correct responses on the new inference questions relative to both the correct answer feedback and no-feedback conditions (.45 vs. .30), $t(38) = 3.13$, $SED = .05$, $d = .90$; and (.45 vs. .28), $t(38) = 2.97$, $SED = .05$, $d = 1.09$, respectively. The correct answer and no-feedback conditions did not differ ($t < 1$). In addition, an item analysis was conducted for the critical comparison between the correct answer and explanation feedback conditions by computing a t test with items as the unit of observation instead of subjects. This item analysis confirmed that explanation feedback produced superior transfer, $t(40) = 2.04$, $SED = .07$, $d = .61$.

Discussion

The results of Experiment 1 showed that the benefits of explanation feedback depend on how learning is assessed. Replicating the findings of previous studies, explanation feedback produced equivalent performance relative to correct answer feedback when retention was assessed with repeated questions on the final test (e.g., Gilman, 1969; Kulhavy et al., 1985; Mandernach, 2005; Pridemore & Klein, 1995; Sassenrath & Gaverick, 1965; Smits et al., 2008; Whyte et al., 1995). However, when the final test assessed understanding by requiring subjects to transfer their knowledge of the concept to a new context, explanation feedback led to better performance than correct answer feedback. If it can be replicated, this novel finding is important because it opens the door to a promising new direction for future research: the use of elaborative feedback to promote transfer of learning.

Experiment 2

One of the goals of Experiment 2 was to replicate the novel finding from Experiment 1 that explanation feedback produced better transfer to new inference questions than did correct answer feedback. A second goal was to investigate a potential explanation for this finding. As described in the introduction, the ability to transfer knowledge to new contexts requires understanding; however, transfer also requires retention, especially if the ability to

transfer knowledge is assessed after a delay, such as in Experiment 1. One way of conceptualizing the process of transfer involves breaking it down into three steps: (1) The learner must *recognize* that previously acquired knowledge is relevant, (2) the learner must *recall* that knowledge, and (3) the learner must *apply* that knowledge to the new context (see Barnett & Ceci, 2002). In this conceptualization, the first two steps in the transfer process reflect retention of knowledge, whereas the third step reflects understanding.

In Experiment 1, the first step (recognition) was unlikely to have been a problem: All subjects were instructed that the final test questions were about information that they had read in the passages, and therefore they recognized that they had to recall and apply their knowledge about the passages. Thus, the difference between the two feedback conditions in the ability to transfer knowledge must have been due to differences in recall, application, or both of these steps. Each explanation feedback message re-presented some information from the passage about the critical concept, so it is possible that this re-presentation boosted later recall of that information on the final test. In contrast, subjects who received correct answer feedback might have been less likely to recall this information because they had only studied it once when they read the passages. Although it is possible that differences in recall (Step 2) may have contributed to the results, we believe it is more likely that explanation feedback fostered a deeper understanding of the concepts, which facilitated the application of that knowledge to complete the final step.

In order to investigate this idea, a second phase was added to the final test in which all subjects reanswered the inference questions with the explanation feedback present (i.e., regardless of whether they had received explanation or correct answer feedback on the initial test). The rationale for the inclusion of the “reanswer” phase was that it would separate the recall and application steps in the transfer process (see Table 1 for a schematic explanation of the logic). As described above, any difference in performance between the correct answer and explanation feedback conditions when answering the new inference questions could be due to recall, application, or both of these components. By allowing subjects to consult the explanation feedback during the subsequent reanswer phase, the need to retain the information would be eliminated. Thus, any difference in performance between the two feedback conditions in the reanswer phase would reflect the subjects’ ability to apply their knowledge (i.e., their depth of understanding).

In addition to the inclusion of the reanswer phase, a few other changes were made for Experiment 2. First, the type of feedback variable was manipulated within subjects in order to show that this finding would generalize across experimental designs. Second, the no-feedback condition was dropped in order to maximize the number of items in the explanation and correct answer feedback conditions. Third, the final test consisted of only new inference questions (i.e., no repeated questions) in order to focus on replicating the key finding from Experiment 1.

Method

Participants. Twenty-four Duke University students participated for either course credit or a payment. One additional subject was excluded for not following the instructions.

Design. A single variable (type of feedback: correct answer, explanation) was manipulated within subjects, between materials.

Materials. The materials from Experiment 1 were used again.

Procedure. The procedure was the same as Experiment 1 except for the following changes. First, subjects received correct answer feedback on 10 of the definition questions on the initial test and explanation feedback for the other 10 questions. Second, the final test consisted of 20 new inference questions (no questions were repeated from the initial test). Third, the final test consisted of two phases. In Phase 1, subjects answered the new inference questions in the same manner as Experiment 1. In Phase 2, they were given the opportunity to reanswer each inference question while also viewing the relevant explanation feedback (i.e., regardless of whether they had seen the explanation feedback on the initial test or not). Subjects were told that they could re-enter their initial response or modify their response based on the information presented in the explanation feedback.

Results

Scoring. Again, two coders independently scored the responses. Reliability was almost perfect ($\kappa = .98$), and the first author (ACB) resolved the few disagreements.

Initial test performance. Overall, subjects correctly answered a little less than half the questions (grand $M = .44$), and there was no significant difference between the two feedback conditions ($t < 1$).

Final test performance. The left panel of Figure 2 shows the proportion of correct responses on the initial answer phase of the final test as a function of feedback condition on the initial test.

Table 1
The Logic Behind the Two-Phase Final Test Used in Experiment 2

Step in the transfer process	Final test phase	
	Initial answer to new inference question	Reanswer with explanation feedback
Recognition	EX = CA	EX = CA
Recall	EX > CA	EX = CA
Application	EX > CA	EX > CA

Note. When initially answering the new inference questions, there should be no difference between the two feedback conditions with respect to the recognition component of the feedback process; however, the explanation feedback condition could lead to better recall and/or application. In the reanswer phase, both recognition and recall are equated; thus, the superiority of explanation feedback over correct answer feedback must be due to the application component. EX = explanation feedback; CA = correct answer feedback.

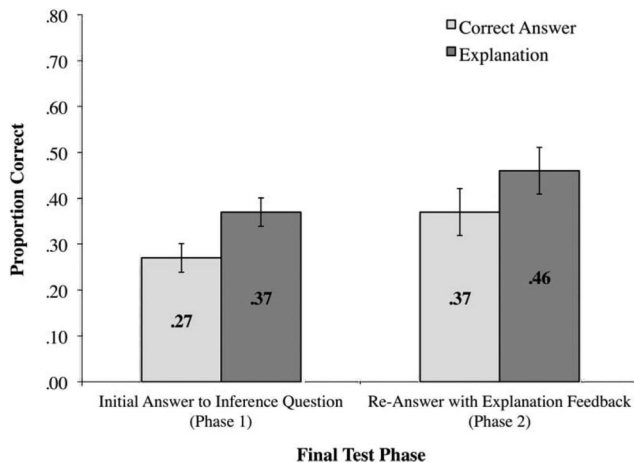


Figure 2. Proportion of correct responses on the final test as a function of feedback condition on the initial test in the initial answer (left side) and reanswer (right side) phases of the final test in Experiment 2. Error bars represent 95% confidence intervals.

Replicating the key result from Experiment 1, explanation feedback led to a significantly greater proportion of correct responses on the new inference questions relative to correct answer feedback (.37 vs. .27), $t(23) = 4.18$, $SED = .02$, $d = .50$.

The right panel of Figure 2 shows the proportion of correct responses on the reanswer phase of final test as a function of feedback condition on the initial test. Overall, the opportunity to reanswer the inference questions with the explanation feedback present improved in both the explanation feedback and correct answer feedback conditions; however, explanation feedback still produced a significantly greater proportion of correct responses than correct answer feedback (.46 vs. .37), $t(23) = 2.64$, $SED = .04$, $d = .37$. In order to compare performance on the two phases, a 2 (final test phase: initial answer, reanswer) \times 2 (type of feedback: correct answer, explanation) ANOVA was conducted. This analysis revealed significant main effects of final test phase, $F(1, 23) = 14.50$, $MSE = .02$, $\eta^2 = .39$, and type of feedback, $F(1, 23) = 15.86$, $MSE = .01$, $\eta^2 = .41$. However, the interaction was not significant ($F < 1$). In addition, an item analysis was conducted by computing the same 2 \times 2 ANOVA with items as the unit of observation instead of subjects. This item analysis revealed the same pattern of results: significant main effects of phase, $F(1, 19) = 18.45$, $MSE = .01$, $\eta^2 = .15$, and type of feedback, $F(1, 19) = 5.98$, $MSE = .01$, $\eta^2 = .15$, but no significant interaction ($F < 1$).

Discussion

Experiment 2 replicated the key novel finding from Experiment 1. When subjects received explanation feedback on the initial test, they were more successful at transferring their knowledge on the new inferences questions than when they received correct answer feedback. The additional information contained in the explanation feedback message fostered better understanding of the critical concepts, which enabled subjects to apply this knowledge to answer new inference questions. Importantly, this result also shows that the effect generalizes across experimental design—type of

feedback was manipulated between subjects in Experiment 1 and within subjects in Experiment 2.

When subjects had the opportunity to reanswer the inference questions with the explanation feedback present, the results were intriguing. Performance improved in the explanation feedback condition, which suggests that some of the information from the feedback had been forgotten; once subjects were re-presented with this information, they were able to successfully apply this knowledge to answer the inference questions. Performance also improved in the correct answer feedback condition. This improvement presumably also reflects the recall component of the transfer process—because subjects did not receive the explanation feedback on the initial test, they may not have retained this information (unless they remembered it from the passage). Practically speaking, this finding is important because it shows that giving explanation feedback after a delay can still help to improve transfer, which is consistent with recent research that shows a benefit of feedback even when its presentation is delayed (e.g., A. C. Butler, Karpicke, & Roediger, 2007; Metcalfe, Kornell, & Finn, 2009).

Most importantly, the difference in performance between the two feedback conditions for the initial answers to the new inference questions was also observed in the subsequent reanswer phase. In both feedback conditions, the presence of the explanation feedback while reanswering the inference questions meant that the burden to recall this information was removed, and any difference between the two conditions had to be due to their ability to apply their knowledge. Receiving explanation feedback on the initial test may have enabled subjects to acquire a deeper understanding of the critical concepts, which helped them to correctly answer more inference questions in the reanswer phase. Furthermore, this finding suggests that it may be particularly important to receive the explanation feedback soon after retrieving a concept from memory because the difference between the two feedback conditions persisted in the reanswer phase when the explanation feedback was always present. We turn now to discussing the importance of these findings in the context of the broader feedback literature.

General Discussion

The present research helps to resolve a paradox about elaborative feedback. Although elaborative feedback is assumed to benefit learners and it is often included in instructional methods (e.g., Corbett et al., 1997; Gibbons & Fairweather, 1998), reviewers of the feedback literature had concluded that increasing the complexity of the feedback message does not benefit learning (e.g., Bangert-Drowns et al., 1991; Kulhavy & Stock, 1989). With respect to the existing evidence in the literature, this conclusion was warranted—many studies that have isolated the effects of greater feedback complexity have found no benefit of elaborative feedback relative to correct answer feedback (e.g., Andre & Thieman, 1988; Gilman, 1969; Kulhavy et al., 1985; Peeck, 1979; Pridemore & Klein, 1995; Sassenrath & Gaverick, 1965; Whyte et al., 1995). However, all of these studies assessed retention of the correct response to a previously presented question rather than deeper understanding of the material. When understanding was assessed in the present study, explanation feedback produced better performance than correct answer feedback. This finding suggests the need for a fundamental reevaluation of how elaborative feedback affects learning.

Why did explanation feedback produce superior performance on new inference questions relative to correct answer feedback? One might expect to find an answer to this question among the various theories that have been proposed to explain how feedback affects learning. However, many of these theories do not address this question at all because they seek to describe the effects of feedback at a more complex level than that of a single task (e.g., D. L. Butler & Winne, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Such “macrolevel” theories model the influence of feedback on various student behaviors, such as self-regulation, learning strategies, and motivation, during a continuous process of learning that includes repeated presentations of feedback. Although other theories provide a “microlevel” account of learning from feedback during a single task, these theories are either too general (e.g., Bangert-Drowns et al., 1991) or focus on explaining other feedback phenomena (e.g., the relationship between response confidence and feedback processing; Kulhavy, 1977). Kulhavy and Stock (1989) put forth the only theoretical framework that specifically addresses the effects of elaborating the feedback message beyond providing the correct answer. Despite their efforts to develop a coherent account of how elaboration affects learning, they were “unable to reach any useful conclusion regarding how the elaborative component of the feedback operates” (Kulhavy & Stock, 1989, p. 289). Recent microlevel reviews of the feedback literature describe many of these theories but offer no new ideas regarding elaborative feedback (e.g., Mory, 2004; Shute, 2008).

Given the dearth of existing feedback theory upon which to draw, we looked to theories in other domains in order to develop an explanation for our findings. One relevant theory is the framework proposed by Barnett and Ceci (2002) to explain the process of transfer and the factors that influence whether it will occur. As described above, they conceptualize the process of transfer in terms of three steps: recognition, recall, and application. Both correct answer and explanation feedback can improve the retention of specific knowledge, which would facilitate later recall of the information (i.e., the second step in the transfer process); this conclusion is supported by the finding that the two types of feedback produced equivalent performance on the definition questions that were repeated on the final test in Experiment 1. However, explanation feedback may also enable learners to better comprehend the concepts, thus facilitating the application of that knowledge to new contexts (i.e., the third step in the transfer process). The results of the reanswer phase in Experiment 2 support this conclusion. When subjects reanswered the inference questions with the explanation feedback present, the superiority of explanation feedback persisted even though the recall demands were removed, suggesting that the locus of the effect is the application step of the transfer process.

Another way of framing our findings is through the lens of text-processing theories that conceptualize the development of understanding as a process that requires representing a text on multiple levels (for a review, see Graesser, Millis, & Zwaan, 1997; Kintsch, 1998). Such theories often differentiate between three levels of representation: the *surface level*—the specific words and syntax used in the text; the *textbase*—an abstract representation of the ideas and their connections; and the *situation model*—a personal interpretation of the text that often includes preexisting knowledge. According to most theories, the situation model is the representational level that reflects deep understanding and sup-

ports the transfer of knowledge. Within the context of the present study, processing the explanation feedback after an initial retrieval attempt may have helped subjects to improve their situation model of the text and achieve a deeper understanding. A more developed situation model would be expected to enable superior transfer of knowledge to the new inference questions, which were aligned with this representational level. In contrast, the repeated questions used to assess retention were aligned with memory for the textbase, and thus explanation feedback would not be expected to benefit performance on these items relative to correct answer feedback.

One remaining puzzle is why explanation feedback was effective at facilitating understanding when it was given on the initial test, but it did not have the same effect on the correct answer condition when it was presented during the reanswer phase of the final test. Although additional research will be needed to further explore this finding, one potential explanation revolves around the concept of memory reconsolidation. In general, practice retrieving the critical concepts from memory would be expected to help subjects to better retain these concepts and transfer them to new contexts, regardless of feedback condition (e.g., A. C. Butler, 2010; Roediger & Butler, 2011). However, retrieval may also reopen a memory so that it must be reconsolidated, meaning that the memory enters a labile state in which it can be altered (e.g., Hupbach, Gomez, Hardt, & Nadel, 2007; for a review, see Dudai, 2006; Lee, 2009). For example, a recent study by Finn and Roediger (2011) showed that postretrieval processing of new information results in the integration of this information into the existing memory, thereby enhancing retention. In the present study, postretrieval processing of the explanation feedback on the initial test may have resulted in the information being integrated into the memory of the concept, thus building a deeper understanding (i.e., a more developed situation model). Retrieval during the final test should also involve reopening the memory, giving a chance for both groups of subjects to integrate the explanation feedback presented in the reanswer phase into their memories; however, it may be that the memory must be successfully reconsolidated (over time) before a deeper understanding is developed. Although admittedly somewhat speculative, this reconsolidation hypothesis provides a potential starting point for follow-up studies.

The present findings open the door for new research that investigates the role of feedback in promoting transfer of knowledge. The need for this research is apparent with respect to all types of elaborative feedback, but also more generally with other factors that influence the efficacy of feedback. The vast majority of feedback studies in the literature use final tests with repeated questions to assess retention of knowledge. Although retention is certainly an important learning outcome, so too is understanding. Thus, there is a great need for research on how feedback affects transfer for both theoretical and pedagogical purposes. If understanding is ignored as a learning outcome, many promising methods of providing feedback may be misconceived and overlooked. For example, one method that may help to produce substantial understanding is to give students correct answer feedback and then have them generate their own explanations for why their response is correct or incorrect. Previous studies have not found a benefit of such a procedure relative to simply providing correct answer feedback (e.g., McDaniel & Fisher, 1991); however, these studies have measured retention rather than understanding. In summary, the findings of the present study indicate that transfer of knowl-

edge represents a fruitful new frontier for feedback research—it is time for feedback researchers to move beyond measuring retention and investigate how feedback affects understanding.

References

- Andre, T., & Thieman, A. (1988). Level of adjunct question, type of feedback, and learning concepts by reading. *Contemporary Educational Psychology, 13*, 296–307. doi:10.1016/0361-476X(88)90028-8
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5*, 7–74.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1118–1133. doi:10.1037/a0019902
- 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., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a meta-cognitive error: Feedback enhances retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918–928. doi:10.1037/0278-7393.34.4.918
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Psychology, 65*, 245–281.
- Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 849–874). New York, NY: Elsevier.
- Dudai, Y. (2006). Reconsolidation: The advantage of being refocused. *Current Opinion in Neurobiology, 16*, 174–178. doi:10.1016/j.conb.2006.03.010
- Fazio, L. K., Huelsner, B. J., Johnson, A., & Marsh, E. J. (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18*, 335–350. doi:10.1080/09658211003652491
- Finn, B., & Roediger, H. L., III. (2011). Enhancing retention through reconsolidation: Negative emotional arousal following retrieval enhances later memory. *Psychological Science, 22*, 781–786. doi:10.1177/0956797611407932
- Gibbons, A. S., & Fairweather, P. G. (1998). *Computer-based instruction: Design and development*. Englewood Cliffs, NJ: Educational Technology.
- Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology, 60*, 503–508. doi:10.1037/h0028501
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). Auto-Tutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*, 612–618. doi:10.1109/TE.2005.856149
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163–189. doi:10.1146/annurev.psych.48.1.163
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. doi:10.3102/003465430298487
- Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory, 14*, 47–53. doi:10.1101/lm.365707
- Jarvis, B. G. (2004a). Medialab (Version 2004.2.87) [Computer software]. New York, NY: Empirisoft Corporation.
- Jarvis, B. G. (2004b). DirectRT (Version 2004.1.0.55) [Computer software]. New York, NY: Empirisoft Corporation.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi:10.1037/0033-2909.119.2.254
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211–232.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certainty. *Educational Psychology Review, 1*, 279–308. doi:10.1007/BF01320096
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*, 285–291. doi:10.1016/0361-476X(85)90025-6
- Lee, J. L. C. (2009). Reconsolidation: Maintaining memory relevance. *Trends in Neurosciences, 32*, 413–420. doi:10.1016/j.tins.2009.05.002
- Mandernach, B. J. (2005). Relative effectiveness of computer-based and human feedback for enhancing student learning. *The Journal of Educators Online, 2*, 1–17.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192–201. doi:10.1016/0361-476X(91)90037-L
- McGeoch, J. A. (1942). *The psychology of human learning: An introduction*. New York, NY: Longmans, Green and Co. doi:10.2307/2262568
- Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adult's vocabulary learning. *Memory & Cognition, 37*, 1077–1087. doi:10.3758/MC.37.8.1077
- 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.
- 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
- Peock, J. (1979). Effects of differential feedback on the answering of two types of questions by fifth- and sixth-graders. *British Journal of Educational Psychology, 49*, 87–92. doi:10.1111/j.2044-8279.1979.tb02401.x
- Phye, G. D., & Sanders, C. E. (1994). Advice and feedback: Elements of practice for problem solving. *Contemporary Educational Psychology, 19*, 286–301. doi:10.1006/ceps.1994.1022
- Plowman, L., & Stroud, J. B. (1942). Effect of informing pupils of the correctness of their responses to objective test questions. *Journal of Educational Research, 36*, 16–20.
- Pridmore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-assisted instruction. *Contemporary Educational Psychology, 20*, 444–450. doi:10.1006/ceps.1995.1030
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20–27. doi:10.1016/j.tics.2010.09.003
- Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology, 14*, 43–49.
- Sassenrath, J. M., & Gaverick, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology, 56*, 259–263. doi:10.1037/h0022474
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189. doi:10.3102/0034654307313795
- Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: Effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*, 183–193. doi:10.1080/10494820701365952

Travers, R. M. W., Van Wagenen, R. K., Haygood, D. H., & McCormick, M. (1964). Learning as a consequence of the learner's task involvement under different conditions of feedback. *Journal of Educational Psychology*, 55, 167–173. doi:10.1037/h0048319

Whyte, M. M., Karolick, D. M., Neilsen, M. C., Elder, G. D., & Hawley, W. T. (1995). Cognitive styles and feedback in computer-assisted instruction. *Journal of Educational Computing Research*, 12, 195–203. doi:10.2190/M2AV-GEHE-CM9G-J9P7

Appendix A

Sample Passages Used in Experiments 1 and 2

These passages are associated with the sample questions and feedback provided in Table 1.

The Respiratory System

Humans breathe in and out anywhere from 15 to 25 times per minute. The main function of the respiratory system is gas exchange between the external environment and the circulatory system. A gas that the body needs to get rid of, carbon dioxide, is exchanged for a gas that the body can use, oxygen. The lungs are the most critical component of the respiratory system because they are responsible for the oxygenation of the blood and the concomitant removal of carbon dioxide from the circulatory system. Gas exchange occurs in tiny, thin-walled air sacs called alveoli, which lie at the end of the many branches of tubes in the lungs. Within the alveoli, gas exchange occurs as a result of diffusion. Diffusion is the movement of particles from a region of high concentration to a region of low concentration. The oxygen concentration is high in the alveoli, so oxygen diffuses across the alveolar membrane into the pulmonary capillaries, which are small blood vessels that surround each alveolus. The hemoglobin in the red blood cells passing through the pulmonary capillaries has carbon dioxide bound to it and very little oxygen. The oxygen binds to hemoglobin and the carbon dioxide is released. Since the concentration of carbon dioxide is high in the pulmonary capillaries relative to the alveolus, carbon dioxide diffuses across the alveolar membrane in the opposite direction. The exchange of gases across the alveolar membrane occurs rapidly—usually in fractions of a second.

Humans do not have to think about breathing because the body's autonomic nervous system controls it. The respiratory centers that control the rate of breathing are located in the pons and medulla oblongata, which are both part of the brainstem. The neurons that live within these centers automatically send signals to the diaphragm and intercostal muscles to contract and relax at regular intervals. Neurons in the cerebral cortex can also voluntarily influence the activity of the respiratory centers. A region within the cerebral cortex, called motor cortex, controls all voluntary motor functions, including telling the respiratory center to speed up, slow down, or even stop. However, the influence of the nerve centers that control voluntary movements can be overridden by the autonomic nervous system. Several factors can trigger such an override. One of these factors is the concentration of oxygen in the

blood. Specialized nerve cells within the aorta and carotid arteries called peripheral chemoreceptors monitor the oxygen concentration of the blood. If the oxygen concentration decreases, the chemoreceptors signal to the respiratory centers in the brain to increase the rate and depth of breathing. These peripheral chemoreceptors also monitor the carbon dioxide concentration in the blood. Another factor is chemical irritants. Nerve cells in the airways can sense the presence of unwanted substances like pollen, dust, water, or cigarette smoke. If chemical irritants are detected, these cells signal the respiratory centers to contract the respiratory muscles, and the coughing that results expels the irritant from the lungs.

Vaccines

A vaccine is a biological preparation that establishes or improves immunity to a particular disease. Most vaccines are prophylactic, which means that they prevent or ameliorate the effects of a future infection by any natural pathogen. However, vaccines have also been used for therapeutic purposes, such as for alleviating the suffering of people already afflicted with a disease. The early vaccines were inspired by the concept of variolation, which originated in Asia during the 13th century. Variolation is a technique in which a person is deliberately infected with a weak form of a disease by inhaling it through the nose or mouth. Upon recovery, the individual was immune to the disease. A small proportion of the people who were variolated died, but nowhere near the proportion that died when they contracted the disease naturally. By the 18th century, knowledge of variolation had spread to Europe where medical researchers Edward Jenner and Louis Pasteur transformed the ancient technique into the modern day practice of inoculation with vaccines. Inoculation represented a major breakthrough because it reduced the risk of vaccination, while maintaining its effectiveness. Inoculation is the practice of deliberate infection through a skin wound. This new technique produces a smaller, more localized infection relative to variolation in which inhalation of viral particles spreads the infection more widely. The smaller infection works better because it is sufficient to stimulate immunity to the virus, but it keeps the virus from replicating enough to reach levels of infection likely to kill a patient.

(Appendices continue)

Vaccines work because they prepare the immune system to deal with pathogens that it may encounter in the future. When a vaccine is given, the immune system recognizes the vaccine agents as foreign, destroys them, and then “remembers” them. When the real version of virus comes along, the body recognizes it and destroys the infected cells before they multiply. Of course, vaccines do not guarantee complete protection against the disease because sometimes a person’s immune system does not respond for various reasons. Still, even when a vaccinated individual does develop the disease vaccinated against, the disease is likely to be milder than without vaccination. Overall, the invention of vaccines has led to a marked decrease in the prevalence of deadly diseases, such as smallpox, polio, measles,

and typhoid. As long as the vast majority of people are vaccinated, it is much more difficult for an outbreak of disease to occur and spread because of herd immunity. Herd immunity describes a type of immunity that occurs when the vaccination of a portion of the population (or herd) provides protection to unvaccinated individuals. Herd immunity theory proposes that for diseases passed from person-to-person, it is more difficult to maintain a chain of infection when large numbers of a population are immune. The higher the proportion of individuals who are immune, the lower the likelihood that a susceptible person will come into contact with an infected individual. Despite potential protection from herd immunity, mainstream medical opinion is that everyone should be vaccinated.

Appendix B

Sample Questions and Feedback Taken From Passages on the Respiratory System and Vaccines, Respectively

Retention questions were used on the initial test and repeated on the final test, whereas transfer questions were only used on the final test.

The Respiratory System

Retention Question: What is the process by which gas exchange occurs in the part of the human respiratory system called the alveoli?

Correct Answer Feedback: *Gas exchange occurs within the alveoli through diffusion.*

Explanation Feedback: *Gas exchange occurs within the alveoli through diffusion. Diffusion is the movement of particles from a region of high concentration to a region of low concentration. The oxygen concentration is high in the alveoli and the carbon dioxide concentration is high in the pulmonary capillaries, so the two gases diffuse across the alveolar membrane in opposite directions towards lower concentrations.*

Inference Question: If people are having trouble breathing, they are often given pure oxygen to inhale. How does breathing pure oxygen facilitate gas exchange relative to regular air?

Answer: *Breathing pure oxygen increases the oxygen concentration in the alveoli, so oxygen will diffuse more rapidly across the alveolar membrane into blood in the pulmonary capillaries.*

Vaccines

Retention Question: What vaccination technique did Edward Jenner and Louis Pasteur develop that improved upon the ancient practice of variolation?

Correct Answer Feedback: *Edward Jenner and Louis Pasteur developed the technique of inoculation to improve upon the ancient practice of variolation.*

Explanation Feedback: *Edward Jenner and Louis Pasteur developed the technique of inoculation to improve upon the ancient practice of variolation. Inoculation is the practice of deliberate infection through a skin wound, whereas variolation involves inhaling a weak form of the disease. The new technique produces a smaller, more localized infection that is adequate to stimulate immunity to the virus, but keeps it from replicating enough to be dangerous.*

Inference Question: The recently developed nasal spray flu vaccine, which is inhaled through the nose, contains weakened viruses that only cause infection at the cooler temperatures found within the nose. In what sense does this new method of vaccination combine the techniques of inoculation and variolation?

Answer: *The nasal spray flu vaccine is similar to inoculation in that it produces a smaller, more localized infection, but also like variolation in that the virus is inhaled.*

Received May 16, 2011

Revision received September 24, 2012

Accepted October 11, 2012 ■