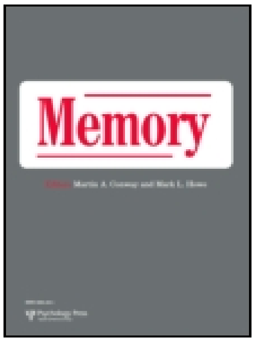

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Using verification feedback to correct errors made on a multiple-choice test

Elizabeth J. Marsh , Jeffrey P. Lozito , Sharda Umanath , Elizabeth L. Bjork & Robert A. Bjork

MEMORY, 2012, 20 (6), 645–653
Psychology Press
Taylor & Francis Group

BRIEF REPORT

Using verification feedback to correct errors made on a multiple-choice test

Elizabeth J. Marsh[1], Jeffrey P. Lozito[2], Sharda Umanath[1], Elizabeth L. Bjork[3], and Robert A. Bjork[3]

[1]Department of Psychology & Neuroscience, Duke University, Durham, NC, USA
[2]SA Technologies, Inc., Atlanta, GA, USA
[3]Department of Psychology, University of California, Los Angeles, CA, USA

A key educational challenge is how to correct students' errors and misconceptions so that they do not persist. Simply labelling an answer as correct or incorrect on a short-answer test (verification feedback) does not improve performance on later tests; error correction requires receiving answer feedback. We explored the generality of this conclusion and whether the effectiveness of verification feedback depends on the type of test with which it is paired. We argue that, unlike for short-answer tests, learning whether one's multiple-choice selection is correct or incorrect should help participants narrow down the possible answers and identify specific lures as false. To test this proposition we asked participants to answer a series of general knowledge multiple-choice questions. They received no feedback, answer feedback, or verification feedback, and then took a short-answer test immediately and two days later. Verification feedback was just as effective as answer feedback for maintaining correct answers. Importantly, verification feedback allowed learners to correct more of their errors than did no feedback, although it was not as effective as answer feedback. Overall, verification feedback conveyed information to the learner, which has both practical and theoretical implications.

Keywords: Feedback; Error correction; Multiple-choice testing.

Our knowledge about the world is imperfect, often leading us to make errors, both during new learning and in retrieving stored knowledge. A key educational challenge is how and when to provide feedback, with the goal of preventing errors from propagating to a later time. Surprisingly, the answer is more complicated than ''give feedback''. Sometimes feedback can be detrimental to later performance, with effect sizes in one meta-analysis ranging from -1.0 to $+1.5$ (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991).

The effectiveness of feedback depends on multiple factors. It is less useful when participants can peek at feedback before answering a question (e.g., Anderson, Kulhavy, & Andre, 1971, 1972), and more useful with a delay between answering a question and receiving feedback (e.g., Butler, Karpicke, & Roediger, 2007). Feedback is more helpful for correcting errors than maintaining correct responses, and providing feedback after correct answers can be counterproductive when time is limited, as the time could be better spent on other activities (Hays, Kornell, & Bjork, 2010).

Address correspondence to: Elizabeth J. Marsh, Department of Psychology & Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708-0086, USA. E-mail: emarsh@psych.duke.edu

A collaborative activity award from the James S. McDonnell Foundation supported this work. The opinions expressed are those of the authors and do not represent the views of the Foundation. A National Science Foundation Graduate Research Fellowship supported the third author. We thank Aaron Johnson for collecting the data and for assistance with manuscript preparation.

http://www.psypress.com/memory http://dx.doi.org/10.1080/09658211.2012.684882

Our focus is on the type of feedback provided, contrasting verification and answer feedback. *Verification feedback*—also called partial, knowledge-of-results, or right/wrong feedback—labels a response as correct or incorrect, whereas *answer feedback* provides the correct answer. Pashler, Cepeda, Wixted, and Rohrer (2005) conducted one of the most thorough experiments evaluating answer and verification feedback. Participants learned English translations of Luganda words (e.g., *leero* means *today*) and took an initial test (e.g., *leero –?*). Critically, participants received no feedback, verification feedback (correct/incorrect), or answer feedback (the correct translation). Participants took a second test immediately and a third test a week later. Receiving answer feedback led to the best performance on both tests. Errors were only corrected after answer feedback; verification feedback was no more effective than receiving no feedback (Pashler et al., 2005), consistent with other work supporting the use of answer feedback over verification feedback (e.g., Bangert-Drowns et al., 1991; Fazio, Huelser, Johnson, & Marsh, 2010; Gilman, 1969; Roberts & Park, 1984; Roper, 1977; Shute, 2008; Waldrop, Justen, & Adams, 1986).

Notably, Pashler et al. (2005) used cued recall: Participants read each Luganda word and recalled its English translation. Educators, however, often use other types of tests, especially multiple-choice exams, which are easy to grade and are typically perceived as more objective. A complete evaluation of verification feedback must consider such tests. Knowing that a multiple-choice selection is correct or incorrect gives the learner more information than knowing if an open-ended response is correct or incorrect. When an open-ended response is labelled as incorrect, there remain an infinite number of alternatives, whereas knowing a multiple-choice selection is incorrect allows the learner to constrain the possibly correct alternatives. In this sense, verification feedback is similar to answer-until-correct feedback, where participants continue selecting multiple-choice answers until they answer correctly. In both cases, the learner can meaningfully winnow down the choices, but with verification feedback, the learner receives no additional feedback about further selections. Answer-until-correct feedback seems to be as effective as answer feedback for error correction (Brosvic, Dihoff, Epstein, & Cook, 2006; Brosvic, Epstein, Cook, & Dihoff, 2005; Butler et al., 2007), suggesting that verification feedback following multiple-choice selections might also be beneficial.

Previous work, however, yields inconsistent conclusions about verification feedback following multiple-choice selections. In some cases, feedback form does not matter (e.g., Lee, 1985; Mandernach, 2005; Mory, 1994), whereas in other cases, verification feedback improves later performance more than receiving answer feedback (Hanna, 1976; Wentling, 1973). Methodological issues likely contributed to these mixed results. For example, when participants learn to criterion (as in Gilman, 1969), differences between feedback conditions are likely to be obscured. Furthermore it is difficult to draw strong conclusions when factors are outside experimental control, such as when participants control how long feedback appears (e.g., Stock, Kulhavey, Pridemore, & Krug, 1992) or the study occurs in a classroom (e.g., Moore & Smith, 1964; Wentling, 1973). Finally, and crucially, the multiple-choice options must be visible when verification feedback appears; otherwise, its usefulness will depend on the ability to remember the multiple-choice alternatives (as in Hancock, Stock, & Kulhavey, 1992).

Theoretically the *desirable difficulties* framework allows predictions about when a learner might benefit from verification feedback. This framework proposes that providing challenges to the learner (e.g., testing, spaced learning) engages processes that enhance subsequent retention and transfer, given that the learner can still succeed (Bjork, 1994; Schmidt & Bjork, 1992). Verification feedback poses a difficulty when the learner makes an error, as it does not provide the correct answer. As will be discussed below, the question is whether this lack of knowledge is always undesirable (as Pashler et al., 2005 suggest) or if it can be desirable.

Some prior research on multiple-choice testing and verification feedback is consistent with the desirable difficulties framework (e.g., Goodman, Wood, & Hendrickx, 2004). The most relevant study is one by Hanna (1976), which involved fifth- and sixth-grade students. The students, who were classified into three ability levels based on a pretest, answered multiple-choice questions about science, mathematics, and social studies. Students received no feedback, verification feedback, or answer-until-correct feedback, with students from the three ability levels equally dispersed across feedback conditions. When the students were later tested, both ability level and prior feedback condition affected performance. Verification feedback led to the best performance for the highest-ability students, whereas answer-until-correct feedback was best for lowest-ability students.

Hanna explained these results as owing to the "stimulation of meaningful discovery" (1976, p. 205) for the highest-achieving students, who presumably deduced the correct answers after learning which ones were wrong. Thus the learner may benefit from discovering the correct answer (Gilman, 1969), similar to benefits of generation in standard episodic memory tasks (Jacoby, 1978; Slamecka & Graf, 1978). Consistent with the desirable-difficulties framework, however, verification feedback will not be desirable if students cannot overcome the difficulty of deducing the correct answer following an error (as in Hanna's case, where the lowest-ability students did not benefit from verification feedback).

We evaluated these ideas in an experiment where the likelihood of discovering the correct answer was directly manipulated, without depending on pre-existing ability differences among individuals. To achieve this goal, we manipulated the number of alternatives available on each multiple-choice question (two, three, or four). Verification feedback provides more information when paired with fewer alternatives; with two alternatives, it should be functionally equivalent to answer feedback. We also varied whether participants were required explicitly to make a second choice after each *error* on verification feedback trials. If participants naturally make second-choice responses covertly, this manipulation should have no effect. If, however, participants need to be prompted to make another choice, the benefits of verification feedback should be greater in the explicit re-answer condition.

We compared verification feedback to both answer feedback and a no-feedback control (manipulated within-participants). The predictions differ depending on whether multiple-choice selections are correct. Given a correct selection, verification and answer feedback should provide equivalent information, with both promoting better performance later than no feedback (as both types of feedback are known to help students retain correct guesses; Butler, Karpicke, & Roediger, 2008; Fazio et al., 2010). For error correction, the predictions depend on the number of multiple-choice alternatives. Of particular interest is the two-alternative condition, where answer and verification feedback are functionally equivalent, but the learner either is told the correct answer or infers it.

Additionally, we varied the timing of the final short-answer test, because desirable difficulties often yield benefits on delayed but not immediate tests. For instance, on immediate tests an individual may benefit from restudying, but on delayed tests, there is often a clear advantage of retrieval practice over study (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006). The critical question is whether the benefits of deducing the correct answer (after an error, as required by verification feedback) are greater on a delayed test. The answer to this question should depend on students' ability to deduce the answer, which will be more likely with fewer multiple-choice alternatives.

In summary, to examine any memorial benefits of receiving verification feedback following multiple-choice selections, we manipulated feedback condition, number of multiple-choice alternatives, requirement to re-answer explicitly following an error on verification feedback trials, and delay until the final test. Of critical interest are the effects of these factors on participants' later ability to correct errors and maintain correct answers.

METHOD

Participants

A total of 48 Duke University undergraduates participated for monetary compensation.

Design and materials

We used a 4 (Number of Multiple-choice Alternatives: Zero [not tested], Two, Three, Four) × 3 (Feedback: None, Verification, Answer) × 2 (Final Test Timing: Immediate, Delayed) × 2 (Re-answer: Yes, No) mixed design, with number of alternatives, feedback condition, and final test timing manipulated within-participants and the chance to re-answer questions paired with verification feedback manipulated between-participants.

A total of 144 general knowledge questions (with four alternatives) were adapted from Roediger and Marsh (2005). One lure was randomly eliminated from each question to create the three-alternative questions; a second lure was randomly eliminated from each to yield the two-alternative questions. The zero-alternative condition refers to questions not initially tested.

For counterbalancing purposes, the facts were divided into two sets of 72 items. Half of the participants answered set A on the immediate short-answer test and set B on the delayed test; the other half did the opposite. Each set was further divided into three groups of 24 items, with each group rotated through the three feedback conditions across participants. Each 24-item group was further divided into four six-item sub-groups that were rotated through the four multiple-choice alternative conditions across participants. Thus, each participant experienced all 12 combinations of the feedback and multiple-choice conditions, with half of the items re-tested immediately and half after 48 hours.

The initial multiple-choice test consisted of 108 questions (e.g., *What is the capital of Belize?*), equally distributed across the two-, three-, and four-alternative formats. One-third of the questions were paired with no feedback, one-third with verification feedback, and one-third with answer feedback.

The short-answer questions were identical to the multiple-choice questions, except no alternatives were presented. Half of the 144 questions were tested immediately and half after 48 hours, with no feedback provided on either test.

Procedure

In the first session participants answered 108 multiple-choice questions. Before receiving those questions participants learned about all three feedback conditions and were told that the lack of feedback did not imply anything about the correctness of their answers. For each question, participants chose one of the presented alternatives; because the three trial types were intermixed, the participants could not know whether they would receive feedback until after answering each question. On answer-feedback trials, the question and alternatives re-appeared for 5 seconds with an asterisk marking the correct answer. On verification-feedback trials, the question (plus alternatives) re-appeared with either the word "CORRECT" or "INCORRECT" below. In the no re-answer condition, verification feedback appeared for 5 seconds until another question was presented. In the re-answer condition "CORRECT" appeared for 5 seconds following correct answers; "INCORRECT" appeared following errors until a second answer was selected (an average of 2.3 seconds); no feedback was provided on second answers.

A five-minute distractor phase involving Sudoku puzzles followed the initial test. All participants then took the immediate short-answer test, answering 72 questions (participants were not explicitly informed of this test or the delayed test ahead of time). To discourage guessing, participants were instructed to respond "don't know" if they did not know the answer.

After 48 hours, participants took the short-answer test on the remaining 72 questions. The instructions and procedures were identical to those of the immediate test. Upon completion, participants were debriefed and dismissed.

RESULTS

Unless otherwise specified, alpha-levels were set at .05.

Performance on the initial multiple-choice test

As expected, participants correctly answered a greater proportion of multiple-choice questions when selecting from two alternatives (M = .62) as compared to three (M = .54) or four alternatives (M = .45), $F(2, 94) = 62.61$, $MSE = .02$, $\eta_P^2 = .57$. Performance did not differ significantly based on feedback condition ($F < 1$), which was expected since questions were answered before receiving any feedback.

Participants in the re-answer condition had a second chance to answer multiple-choice questions following errors on verification feedback trials. On the second try, these participants answered almost all two-alternative questions correctly (M = .96), as compared to a smaller proportion of three-alternative questions (M = .65) and four-alternative questions (M = .50). Explicit instructions to re-answer had no impact on later short answer performance, $F(1, 46) = 1.06$, $MSE = .12$, so this variable was not included in the analyses that follow.

Performance on the final short-answer tests

Two coders scored each answer; they agreed on 99% of judgements, and a third scorer resolved all discrepancies.

TABLE 1
Proportion of final short-answer questions answered correctly

	No Feedback		*Verification Feedback*		*Answer Feedback*	
	Not tested	*Tested*	*Not tested*	*Tested*	*Not tested*	*Tested*
Immediate Test	.22	40	.17	.58	.19	.70
Delayed Test	.22	.35	.19	.44	.22	.53
M	.22	.37	.18	.51	.21	.61

Proportion of final short-answer questions answered correctly as a function of final test timing (immediate vs delayed), prior multiple-choice testing (not tested, tested), and feedback condition (no feedback, verification, answer).

Benefits of prior testing. Our first analyses were aimed at understanding how prior testing benefited performance on the final test and whether these benefits changed when testing was paired with verification feedback. To answer these questions we computed a 2 (Final Test Timing: Immediate or Delayed) × 2 (Prior Multiple-Choice Testing: Not Tested or Tested) × 3 (Feedback Condition: No Feedback, Verification, or Answer) repeated-measures ANOVA, and the relevant data appear in Table 1. The benefits of prior multiple-choice testing depended on whether and what kind of feedback had been received, as reflected in an interaction between feedback condition and prior multiple-choice testing, $F(2,\ 94) = 29.58$, $MSE = .03$, $\eta_P^2 = .39$. Following testing paired with answer feedback, a greater proportion of final questions were answered correctly ($M = .61$), as compared to items that had not been tested previously ($M = .21$; a testing effect of +.40). The size of this testing effect was more powerful than the one observed following verification feedback (+.33), $t(47) = 2.54$, $SE = .03$. However, having received verification feedback was more useful than testing without feedback (+.15), $t(47) = 5.36$, $SE = .03$.

After a delay, the relationship across feedback conditions stayed the same: Answer feedback yielded the strongest testing effect, followed by verification feedback, with the smallest testing effect in the no-feedback condition. Table 1 reflects a three-way interaction between prior testing, feedback condition, and delay, $F(2,\ 94) = 3.85$, $MSE = .02$, $\eta_P^2 = .08$: Delay did not change performance in the no-feedback condition, $t(47) = 1.28$, $SE = .04$, $p = .21$, whereas delay reduced the benefits of testing in both feedback conditions, $t(47) = 4.41$, $SE = .04$ for verification feedback and $t(47) = 4.62$, $SE = .04$ for answer feedback. Notably, the efficacy of verification feedback did not improve over time, contrary to the desirable difficulties framework which predicts that potential benefits of verification feedback would manifest over time.

Connecting initial multiple-choice and final short-answer test performance. The remaining analyses focus on final test performance based on participants' initial multiple-choice test performance. Our goals were to examine error correction and persistence of correct answers across tests. For these analyses, we investigated the effects of number of prior multiple-choice alternatives, with two main predictions: (a) verification feedback will be more effective for correcting errors when learners chose from fewer alternatives; but (b) equally beneficial for maintaining initially correct responses regardless of the number of alternatives presented.

Error correction. Was verification feedback as useful as answer feedback for correcting errors? To answer this question, we analysed the proportion of final short-answer questions answered correctly after *errors* on the initial multiple-choice test (see Table 2). A total of 29 participants had observations in all cells and were included in the 2 (Final Test Timing) × 3 (Feedback Condition) × 3 (Number of Prior Multiple-Choice Alternatives) ANOVA and follow-up analyses on the proportion of errors made on the initial multiple-choice test that were corrected on the final test. Critically, prior feedback condition mattered, $F(2,\ 56) = 65.92$, $MSE = .14$, $\eta_P^2 = .70$, with participants best able to correct errors after receiving answer feedback ($M = .53$), followed by verification feedback ($M = .28$), $t(28) = 5.58$, $SE = .04$, which in turn yielded more error correction than did no feedback ($M = .08$), $t(28) = 7.73$, $SE = .03$. As expected, delay had no effect on error correction in the no-feedback condition, resulting in an interaction between condition and final test timing, $F(2,\ 56) = 5.19$, $MSE = .08$, $\eta_P^2 = .16$. The relative benefits of feedback type were the same on the immediate and delayed tests: Answer

TABLE 2
Proportion of correct answers on the final test following initial multiple-choice errors

	No Feedback			*Verification Feedback*			*Answer Feedback*		
	Two	*Three*	*Four*	*Two*	*Three*	*Four*	*Two*	*Three*	*Four*
Immediate Test	.14	.06	.04	.45	.36	.22	.70	.65	.54
Delayed Test	.10	.08	.06	.28	.25	.12	.35	.49	.49
M	.12	.07	.05	.36	.31	.17	.52	.57	.51

Proportion of correct answers on the final test following initial multiple-choice errors, as a function of final test timing (immediate, delayed), feedback condition (no feedback, verification, answer), and number of prior multiple-choice alternatives (two, three, four).

feedback was always better than verification feedback, which was always better than receiving no feedback.

However, verification feedback was more helpful in some cases than others. Verification feedback helped students correct a greater proportion of errors made on two-alternative questions ($M = .36$) than on four-alternative questions ($M = .17$), $t(28) = 3.98$, $SE = .05$. In contrast, the number of prior alternatives did not impact error correction in the no-feedback, $t(28) = 1.48$, $SE = .05$, $p = .14$, or answer feedback conditions ($t < 1$), leading to an interaction between feedback condition and number of prior multiple-choice alternatives, $F(4,\ 112) = 2.46$, $MSE = .06$, $\eta^2_P = .08$. Notably, when verification and answer feedback were functionally equivalent (after two-alternative questions), answer feedback still led to more error correction, $t(28) = -3.34$, $SE = .05$. That is, even when verification feedback allowed successful inference of the correct answer, it was less effective than receiving the answer.

Maintenance of correct responses. Verification feedback should provide the necessary information to maintain a correct answer on the final test. To test this prediction, we analysed the proportion of final test questions answered correctly following correct selections on the multiple-choice test in a 2 (Final Test Timing) × 3 (Feedback Condition) × 3 (Number of Prior Multiple-Choice Alternatives) ANOVA and follow-up analyses (see Table 3). Six participants were eliminated because they lacked observations in all cells.

In contrast to no feedback ($M = .61$), feedback (either verification or answer, $Ms = .75$) helped participants reproduce their previously selected correct answers, $F(2,\ 82) = 20.55$, $MSE = .09$, $\eta^2_P = .33$. Moreover, with both types of feedback participants performed better on the final test when they had picked the correct answer out of more lures (i.e., benefits of prior verification and answer feedback increased with the number of prior multiple-choice lures), $F(4,\ 164) = 4.30$, $MSE = .06$, $\eta^2_P = .10$. No other interactions reached significance.

DISCUSSION

Overall, verification feedback was more useful for improving final test performance than no feedback. Without feedback, participants corrected only a small proportion of their errors ($M = .08$), but improved dramatically following verification feedback ($M = .28$). However, performance following verification feedback was still well below

TABLE 3
Proportion of correct answers on the final test following initially correct multiple-choice selections

	No Feedback			*Verification Feedback*			*Answer Feedback*		
	Two	*Three*	*Four*	*Two*	*Three*	*Four*	*Two*	*Three*	*Four*
Immediate Test	.72	.71	.77	.76	.86	.89	.75	.85	.91
Delayed Test	.54	.33	.57	.59	.66	.78	.63	.67	.71
M	.63	.52	.67	.67	.76	.83	.69	.76	.81

Proportion of correct answers on the final test following initially correct multiple-choice selections, as a function of final test timing (immediate, delayed), feedback condition (no feedback, verification, answer) and number of prior multiple-choice alternatives (two, three, four).

the level of error correction observed following answer feedback ($M = .53$). Of critical interest was correction of errors made on two-alternative trials, where equivalent information was provided in answer feedback (through a direct statement of the answer) and verification feedback (by process of elimination). Although the same information was conveyed, participants corrected a smaller proportion of errors following verification feedback ($M = .36$) than answer feedback ($M = .52$). This pattern of data is surprising based on the desirable difficulties framework, which predicted that students would benefit from successfully using verification feedback to figure out how to correct their errors on two-alternative questions. Students did successfully overcome the difficulty of verification feedback for two-alternative questions: Performance in the re-answer condition was almost perfect, so the lower error correction rate following verification feedback was unlikely to be due to inability to determine the correct answer. To explain this pattern of data, we must draw on other research that suggests students often later recall both their original errors and the corrective feedback (Butler, Fazio, & Marsh, 2011; Butterfield & Metcalfe, 2001). One possibility is that students may later struggle to recall which response was the correct one versus their error. Learning the correct answer through verification feedback is very similar to the original process of making a response, and thus it may later be hard to distinguish these two answers (whereas after answer feedback, a response and experimenter-provided feedback would be more discriminable in memory).

Verification feedback was even less useful for correcting errors made on three- and four-alternative multiple-choice questions. Data from the re-answer condition reveal the problem: As the number of possible multiple-choice answers increased, students were less able to deduce the correct answer following an error (dropping from .96 to .65 to .50 following two, three, and four alternatives). In the cases of the three- and four-alternative questions, the difficulties were not desirable ones, as the students could not successfully re-answer these questions. One related question is why re-answering had no benefit on later performance, given that students sometimes ascertained the correct answers. Our preferred interpretation is that students automatically make second choices when they learn they have made errors; however, it is also possible that making a second choice does not help learners.

In contrast, a more positive story emerged when maintenance of correct answers was examined: Both verification and answer feedback helped students maintain correct answers. Participants were more likely to reproduce an initially correct answer on the final tests after verification or answer feedback ($Ms = .75$) than following no feedback ($M = .61$). Previously, Pashler et al. (2005) and Guthrie (1971) found no benefit from feedback following correct answers, but Butler et al. (2007) found that feedback benefited low-confidence correct answers. Although we did not collect confidence ratings, we assume that the lack of a study phase (and forced reliance on prior knowledge) meant that often participants were guessing or selecting multiple-choice answers with low confidence. Thus feedback likely helped the retention of correct answers here because many of them were made with lower confidence. In addition our finding that both verification and answer feedback helped to maintain correct answers is similar to Butler et al.'s (2007) finding of no difference between answer and answer-until-correct feedback. Both are situations where the two types of feedback are functionally equivalent (in terms of confirming correct answers), providing essentially the same information.

We note one methodological issue before turning to the implications of our findings. Verification and no-feedback trials did differ in one additional way: Participants were exposed to questions longer on feedback trials than when they did not receive feedback. In the no re-answer condition the question reappeared for 5 seconds paired with the message "CORRECT" or "INCORRECT", whereas the program advanced to the next question on no-feedback trials. This procedure for no-feedback trials is common (e.g., Pashler et al., 2005; Peeck, Van Den Bosch, & Kreupling, 1985; Sassenrath & Gaverick, 1965; Sturges 1978) and making no-feedback participants wait (e.g., for 5 seconds) does not change conclusions about the effectiveness of feedback (Butler et al., 2007, 2008; Butler & Roediger, 2008). Furthermore, two data points from the present study suggest that, rather than exposure time, the information that feedback conveys is key. First, exposure time (5 seconds) was equated for answer feedback and verification without re-answer trials; yet answer feedback was still considerably more powerful. Second, the re-exposure time was reduced in the verification re-answer condition (as participants only needed

2.3 seconds to re-answer questions), and these participants performed no differently from verification participants who received 5 seconds of feedback.

We end with a discussion of educational implications. When grading multiple-choice tests by hand, providing answer feedback may sometimes take more time, but is necessary to maximise error correction. When multiple-choice exams are graded using a scantron machine, some models give teachers the choice between verification and answer feedback. Whether grading by hand or by machine, it is possible that educators might choose verification feedback, with the belief that students will benefit from having to figure out the correct answers, when in reality they will benefit less than if they received answer feedback. Nevertheless, marking multiple-choice answers as right or wrong is better than providing no feedback at all. Additionally, simply marking correct answers as such is enough to maintain those answers in memory. The present results support that answer feedback is the best feedback, but also show that verification feedback can provide some useful information to the learner if questions are in multiple-choice format.

Manuscript received 18 November 2011
Manuscript accepted 27 March 2012
First published online 29 May 2012

REFERENCES

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology*, *62*, 148–156.

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

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.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Brosvic, G. M., Dihoff, R. E., Epstein, M. L., & Cook, M. J. (2006). Feedback facilitates the acquisition and retention of numerical fact series by elementary school students with mathematics learning disabilities. *The Psychological Record*, *56*, 35–54.

Brosvic, G. M., Epstein, M. L., Cook, M. J., & Dihoff, R. E. (2005). Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: Learning in the classroom. *The Psychological Record*, *55*, 401–418.

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

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., Karpicke, J. D., & Roediger, H. L., III (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory,and Cognition*, *34*, 918–928.

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.

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

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

Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology*, *60*, 503–508.

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology*, *98*, 248–262.

Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior*, *10*, 23–28.

Hancock, T. E., Stock, W. A., & Kulhavy, R. W. (1992). Predicting feedback effects from response-certitude estimates. *Bulletin of the Psychonomic Society*, *30*, 173–176.

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). Costs and benefits of feedback during learning. *Psychonomic Bulletin and Review*, *17*, 797–801.

Hanna, G. S. (1976). Effects of total and partial feedback in multiple-choice testing upon learning. *The Journal of Educational Research*, *69*, 202–205.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, *10*, 562–567.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649–667.

Lee, O. M. (1985). The effect of type of feedback on rule learning in computer based instruction (Doctoral dissertation, Florida State University, 1975). *Dissertation Abstracts International*, *46*, 955A.

Mandernach, B. J. (2005). Relative effectiveness of computer-based and human feedback for enhancing student learning. *The Journal of Educators Online*, *2*, 1–17.

Moore, J. W., & Smith, W. I. (1964). Role of knowledge of results in programmed instruction. *Psychological Reports*, *14*, 407–423.

Mory, E. H. (1994). The use of response certitude in adaptive feedback: Effects on student performance, feedback study time, and efficiency. *Journal of Educational Computing Research*, *11*, 263–290.

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.

Peeck, J., Van den Bosch, A. B., & Kreupling, W. J. (1985). Effects of informative feedback in relation to retention of initial responses. *Contemporary Educational Psychology*, *10*, 303–313.

Roberts, F. C., & Park, O. (1984). Feedback strategies and cognitive styles in computer-based instruction. *Journal of Instructional Psychology*, *11*, 63–74.

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

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.

Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology*, *14*, 43–49.

Sassenrath, J. M., & Garverick, C. M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology*, *56*, 259–263.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experiment Psychology: Human Learning and Memory*, *4*, 592–604.

Stock, W. A., Kulhavy, R. W., Pridemore, D. R., & Krug, D. (1992). Responding to feedback after multiple-choice answers: The influence of response confidence. *The Quarterly Journal of Experimental Psychology*, *45A*, 649–667.

Sturges, P. T. (1978). Delay of informative feedback in computer-assisted testing. *Journal of Educational Psychology*, *70*, 378–387.

Waldrop, P. B., Justen, J. E., III, & Adams, T. M., II (1986). A comparison of three types of feedback in a computer-assisted instruction task. *Educational Technology*, 43–45.

Wentling, T. L. (1973). Mastery vs. nonmastery instruction with varying test item feedback treatments. *Journal of Educational Psychology*, *65*, 50–58.