

White Paper

A Brief Review of the Literature on Test Feedback

by
Lucas J. Schippers, Ph.D.



**Assessment
Technology
Incorporated**

Assessment Technology, Incorporated

6700 E. Speedway Boulevard
Tucson, Arizona 85710

Phone: 520.323.9033 • Fax: 520.323.9139

Copyright © Assessment Technology, Incorporated 2012. All rights reserved.

Copyright © Assessment Technology, Incorporated 2012. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the publisher.

Assessment Technology, Incorporated, Publishers

Tucson, Arizona, U.S.A.

Printed in the United States of America.

V1-022112

A Brief Review of the Literature on Test Feedback

by Lucas J. Schippers, Ph.D.
Assessment Technology, Incorporated

Table of Contents

Table of Contents.....	i
I. Introduction	1
II. Feedback’s Effects on Test Takers’ Learning	1
III. Taxonomy of Feedback.....	1
IV. The Effects of Different Feedback Types.....	2
V. Discussion	3
VI. ATI English Language Arts’ Content Specialists’ Guidelines for Writing Feedback and Conclusions	3
VII. Suggestions for Additional Research.....	4
VIII. References.....	4

This page intentionally left blank.

I. Introduction

It is difficult to imagine how learning of any sort can occur without feedback. From our first hesitant steps to understanding the world around us to the pursuit of abstract content knowledge, our actions are shaped by the feedback previous actions engender. Consequently, it is reasonable to assume that feedback intentionally provided to test takers will have some positive impact on their learning; the questions guiding this review of the literature are:

1. To what extent has feedback been empirically shown to improve learning?
2. What types of feedback have been shown to be most effective?
3. How should Assessment Technology Incorporated (ATI) content specialists write test item feedback?

II. Feedback's Effects on Test Takers' Learning

Summarizing the benefits of feedback, Attali and Powers (2010) state that it “helps learners determine performance expectations, judge their level of understanding, and become aware of misconceptions” (p. 22). Moreover, it may “provide clues about the best approaches for correcting mistakes and improving performance” (ibid.). However, feedback does not seem to operate as reinforcement in behaviorist terms, i.e., it does not compel test takers to seek correct answers or avoid incorrect ones (Butler, Karpicke, & Roediger, 2007).

Bangert-Drowns, Kulik, Kulik, and Morgan (1991) theorize that feedback is effective at promoting test takers' learning (usually operationalized in empirical studies as retention of content as measured through test-posttest cycles) when it promotes mindfulness of the content (see also Lhyle and Kulhavy [1987]). Based on the results of their meta-analysis, Bangert-Drowns et al. find that, on average, feedback “seems to make a positive but small contribution to achievement, raising scores by about one fourth of a standard deviation ($ES = 0.26$, $SE = 0.06$)” (p. 224), comparable to raising achievement scores from the 50th to the 59th percentile.

III. Taxonomy of Feedback

There is significant variation in how feedback can be delivered to test takers. Standard feedback is provided when teachers return graded tests to students, possibly holding a discussion of individual items (Butler et al., 2007). Jaehnig and Miller (2007, pp. 221-222) describe additional feedback types that appear frequently in the literature:

- **Knowledge of response (KR).** KR consists simply of notifying the test taker when a response is correct or incorrect.
- **Knowledge of correct response (KCR).** KCR directs the student to the correct answer when an incorrect answer is chosen. By definition, KCR is limited to use with incorrect answers, where it is frequently paired with KR (e.g., “Incorrect. The correct answer is *reinforcement*.”).
- **Elaboration.** As its name suggests, elaboration provides discussion of why a response is correct or incorrect. It can stand on its own or be paired with KR and KCR (e.g., “Incorrect. The correct answer is *reinforcement*. The child's crying *increased* in frequency, so it was reinforcement.”).

- **Answer until correct (AUC).** Less a distinct feedback type as a modality of multiple-choice testing, AUC provides KR while requiring that students ultimately select the correct answer before moving on to another item. It is usually, although not exclusively, incorporated in computer-based assessment.
- **Delayed.** Feedback is delayed if it is withheld from the test taker for a defined period of time (from several minutes to days). While the research on delayed feedback is voluminous, it will not be summarized here (see the section entitled “Suggestions for Additional Research” for further discussion).

IV. The Effects of Different Feedback Types

Not all feedback is equally effective:

- KR seems to be the least powerful of the feedback types. According to meta-analyses conducted by Bangert-Drowns et al. (1991) and Jaehnig and Miller (2007), KR alone shows almost no effect on achievement, comparable to the absence of feedback.
- Butler et al. (2007) contend that the efficacy of feedback is “largely determined by whether the feedback message contains the correct response” (p. 274). Bangert-Drowns et al. (1991) confirm that KCR is superior to KR in improving achievement, and Dempsey and Driscoll (1989) argue that KCR is more efficiently processed by test takers than elaboration and is no less effective. However, Jaehnig and Miller (2007) and Pridemore and Klein (1993) do not find sufficient evidence to recommend KCR over elaboration.
- Bangert-Drowns et al. (1991) argue that elaboration is as effective as KCR, while Jaehnig and Miller (2007) contend that elaboration is superior (although only when combined with KR or KCR). However, the improvement in student achievement noted with elaboration feedback seems to plateau as the feedback becomes more complex. Kulhavy, White, Topp, Chan, and Adams (1985) note that increasing “feedback bulk” (i.e., making elaboration feedback longer and more sophisticated) does not yield progressively better results. Similarly, Lhyle and Kulhavy (1987) note that experiments involving scrambling the word order of feedback (which they concede is an artificial way to boost the complexity of feedback) yielded mixed results.
- AUC seems effective primarily for test items involving higher-order learning (Clariana & Koul, 2006; Clariana, Ross, & Morrison, 1991), although Jaehnig and Miller (2007) find that it is no better than other types of feedback.

It should be emphasized that research findings debunk the argument rooted in behaviorism that feedback offers positive reinforcement. Pashler, Cepeda, and Wixted (2005) argue that correct answer feedback does not extend what test takers remember a week after the initial test event. Consequently, they argue that “a reasonable strategy...would seem to be this: Whenever the subject makes an error, provide feedback and time to process the feedback, but when the subject responds correctly, proceed to the next trial without delay” (p. 7).

V. Discussion

This review of the literature suggests that, whenever possible, feedback should be provided to the student in the form of elaboration of incorrect answers (elaboration of correct answer feedback can be offered, but it should not be expected to make much difference in how test takers respond to test items). Elaboration should be paired with KR or KCR for best results.

Unfortunately, the literature offers little practical guidance in how best to write elaboration feedback, primarily because few studies quantify in precise terms what is meant by “elaboration.” However, research findings do suggest that brevity is preferable to comprehensiveness when writing elaboration feedback. There is likely a point beyond which elaboration becomes overly complex, providing no benefit to the test taker and thereby wasting the content specialist’s time in its composition.

Where elaboration is impossible (e.g., student is asked to locate a fact in a text) or impractical (e.g., with very young children dealing with complex concepts), KCR or KR should be provided, although KCR will most likely prove more effective on its own than KR.

VI. ATI English Language Arts’ Content Specialists’ Guidelines for Writing Feedback and Conclusions

In concert with ATI Assessment and Instructional Design Director, Scott Cunningham, B.S., the ATI English language arts’ (ELA) content specialists developed the following guidelines for writing test item feedback:

- Write feedback in complete sentences.
- Keep feedback as short as possible.
- Explain why incorrect answers are incorrect.
- State rules or define terms, when applicable (e.g., grammar, parts of speech).
- Use quotes from texts, when applicable.
- Do not give away correct answers in explanations of incorrect answers (it may be impossible to avoid providing hints, but try to keep them as oblique as possible).
- Do not be too beholden to writing feedback with strictly grade-level vocabulary.
- Do not worry about feedback seeming repetitious.
- Do not worry about stating that an incorrect answer is based on information not present in a text, when applicable.
- Do not worry about omitting an explanation for an incorrect answer when no good explanation can be provided. In such cases, simply state that an incorrect answer is incorrect.

It should be clear that ELA’s guidelines for writing feedback accord for the most part with what empirical research would qualify as “best practices.” ELA content specialists attempt to offer elaboration whenever possible, and they strive to keep elaboration straightforward and concise. When elaboration is not feasible, they provide KR instead. At this time, ELA content specialists do not mix feedback types, i.e., they do not provide KR in concert with elaboration.

The literature indicates that student achievement could be improved by offering KCR on its own or with elaboration. It also indicates that ELA should be combining elaboration with other feedback types.

VII. Suggestions for Additional Research

This brief review of the literature on test feedback touched on only a few of the significant issues addressed by researchers. Inquiry in the following areas could provide beneficial calibration to ATI practice vis-à-vis provision of feedback:

- **Timing.** Whether feedback is best presented immediately or after a delay is a major concern in the literature.
- **Test anxiety.** Many studies examine the influence of feedback on test anxiety (see, for example, Arkin and Walts [1983]).

One issue that is all but ignored by researchers is how teachers understand and make use of feedback (see Stough and Emmer [1998] for a rare exception). Since ATI content specialists write feedback with the assumption that it will be used by teachers as well as students, it would be helpful to know more about how teachers conceptualize test feedback.

VIII. References

- Arkin, R. M., & Walts, E. A. (1983). Performance implications of corrective testing. *Journal of Educational Psychology, 75*, 561—571.
- Attali, Y., & Powers, D. (2010). Immediate feedback and opportunity to revise answers to open-ended questions. *Educational and Psychological Measurement, 70*(1), 22—35.
- 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.
- 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.
- Clariana, R. B., & Koul, R. (2006). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. *British Journal of Educational Psychology, 76*, 259-270.
- Clariana, R. B., Ross, S. M., & Morrison, G. R. (1991). The effects of different feedback strategies using computer-administered multiple-choice questions as instruction. *Educational Technology Research and Development, 39*(2), 5—17.
- Dempsey, J. V., & Driscoll, M. P. (1989, March). *The effects of four methods of immediate corrective feedback on retention, discrimination error, and feedback study time in computer-based instruction.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

- Jaehrig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *The Psychological Record, 57*, 219—232.
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective effectiveness. *Contemporary Educational Psychology, 10*, 285—291.
- Lhyle, K. G., & Kulhavy, R. W. (1987). Feedback processing and error correction. *Journal of Educational Psychology, 79*, 320—322.
- Pashler, H., Cepeda, N. J., & Wixted, J. T. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3—8.
- Pridemore, D. R., & Klein, J. D. (1993, January). *Learner control of feedback in a computer lesson*. Paper presented at the Convention of the Association for Educational Communications and Technology Sponsored by the Research and Theory Division, New Orleans, LA.
- Stough, L. M., & Emmer, E. T. (1998). Teachers' emotions and test feedback. *International Journal of Qualitative Studies in Education, 11*, 341—361.

This page intentionally left blank.