DOCUMENT RESUME

ED 347 983 IR 015 720

AUTHOR Clariana, Roy B.; And Others
TITLE The Effects of Different Feedback Strategies Using Computer-Administered Multiple-Choice Questions as Instruction.
PUB DATE Feb 92
NOTE 24p.; In: Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology and Sponsored by the Research and Theory Division; see IR 015 706.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Enrichment; *Feedback; Grade 11; High Schools; *Instructional Design; *Instructional Effectiveness; *Intermode Differences; Multiple Choice Tests; Pretests Posttests

ABSTRACT

The present study investigated the effects of using different forms of material with 100 eleventh grade students enrolled in a 5-week CBI (computer based instruction) summer enrichment program in Memphis, Tennessee. The basic design consisted of two conditions of instructional support (text and questions vs. questions only), two testings (immediate vs. retention), five levels of similarity between lesson and posttest questions, and five feedback conditions: Knowledge of Correct Response (KCR), delayed KCR, Answer Until Correct (AUC), questions only (no feedback), and no questions. Results showed significant benefits of feedback over no-feedback, with AUC becoming more advantageous and delayed feedback less so as lesson-posttest question similarity decreased. Also, with decreased question similarity and the availability of supporting text, overall feedback effects tended to decrease. The results are discussed in terms of the information processing effects of the different feedback forms, a factor that CBI designers often fail to exploit in planning feedback conditions. Sample materials, data tables, and graphs are included. (41 references) (Author/BBM)

Title:

The Effects of Different Feedback Strategies Using Computer-Administered Multiple-Choice Questions as Instruction

Authors:

Roy B. Clariana
Steven M. Ross
Gary R. Morrison

Abstract

The present study investigated the effects of using different forms of material. The basic design consisted of two conditions of instructional support (text and questions vs. questions only), two testings (immediate vs. retention), five levels of similarity between lesson and posttest questions, and five feedback conditions: Knowledge of Correct Response (KCR), delayed KCR, Answer Until Correct (AUC), questions only (no feedback), and no questions. Results showed significant benefits for feedback over no-feedback, with AUC becoming more advantageous and delayed feedback less so as lesson-posttest question similarity decreased. Also, with decreased question similarity and the availability of supporting text, overall feedback effects tended to decrease. The results are discussed in terms of the information processing effects of the different feedback forms, a factor that CBI designers often fail to exploit in planning feedback conditions.

The Effects of Different Feedback Strategies Using Computer-Administered Multiple-Choice Questions as Instruction

The use of feedback is a critically important and often neglected attribute in computer-based instruction (CBI). Feedback promotes learning by providing students with information about their responses. Through its interactive capabilities, CBI increases the range of feedback strategies that can be efficiently achieved. Specifically, when incorporated in multiple-choice testing, three common forms include (a) knowledge of response feedback (KOR or KR), which indicates that the learner's response was correct or incorrect; (b) knowledge of correct response feedback (KCR), which identifies the correct response; and (c) elaborative feedback, which identifies the correct response while providing additional explanations (Merrill, 1985).

As would be expected, these forms of feedback may not be equally effective. Several studies have shown KCR to be superior to KOR, and KOR to be superior to no feedback (Gilman, 1969; Kulhavy, 1977; Travers, Van Wagenen, Haygood, & McCormick, 1964; Waldrop, Justin, & Adams, 1986). However, based on his own research and a meta-analysis of studies, Schimmel (1983; 1986) concluded that this hierarchy of immediate feedback types is not so well established. Evidence also suggests that elaborative forms of feedback often produce no significant improvement over KCR, but require a considerable development and implementation cost (Merrill, 1985, 1987; Spock, 1987). Despite years of research, the types of situations in which different feedback forms tend to operate most effectively are still not understood. Part of the reason may be a failure to account adequately for the influences on results of task and learner characteristics as well as the cognitive (as opposed to behavioral) impact of the different feedback treatments employed (see Hannafin & Rieber, 1989). Kulhavy & Stock (1989) further attribute the lack of understanding of feedback effects to the reinforcement emphasis of the operant conditioning paradigm that predominated research and theory for many years. In their current model, they stress the cognitive implications of feedback effects on information processing, while indicating that systematic research illuminating such effects has been minimal.

Usually, feedback is provided to the learner after one response. However, using CBI, a learner may easily be allowed a second try with an item (Dempsey & Driscoll, 1989; Noonan, 1984) or may be required to continue to respond until the correct answer is selected (Pressey, 1926, 1950). The latter orientation is conventionally labeled answer-until-correct (AUC) feedback. Allowing unassisted multiple response tries has considerable intuitive appeal. AUC may engage learners in additional active processing following errors and also ensures that the last response is a correct one, a principle espoused over half a century ago in the contiguity theory of Edwin Guthrie (1935). Unfortunately, there have been relatively few controlled empirical studies to test this interpretation or whether, in general, allowing one response or requiring many responses to an item is more effective (Dempsey & Driscoll, 1989; Noonan, 1984). On the one hand, providing the correct answer after only one response may "short-circuit" learning (Schimmel, 1986). Alternatively, requiring a learner to answer until correct may be frustrating (Dick & Latta, 1970).

Feedback timing is another variable of interest (Hannafin & Reiber, 1989; Kulhavy & Stock, 1989). Feedback may be provided immediately after the learner's response or it may be delayed for either a set period of time or set number of responses, such as at the end of a test. From a recent meta-analysis, Kulik and Kulik (1988) concluded that immediate feedback was best for most learning situations, but delayed feedback was superior in "test-acquisition" studies, i.e., learning situations in which test questions are used as the instruction. Two interpretations are most commonly used to explain test-acquisition benefits. One is termed the interference perseveration hypothesis (Kulhavy & Anderson, 1972). This view holds that an incorrect response proactively interferes with an immediately provided correct response. Delaying the presentation of feedback allows learners time to forget their initial responses, thereby reducing proactive interference effects. However, if such is the case proactive interference should occur not only in test-acquisition studies but in any manipulation of immediate and delayed feedback. A possible explanation concerns the instructional support other than embedded questions that most lessons provide. Such support may consist, for example, of reading passages, pictures, outlines, overviews, or video-clips. This support may serve to make the the material more memorable and thus more resistant to proactive interference.

The second interpretation is based on the rationale that delayed feedback repeats the item presentation at the end of the lesson, thereby providing twice as much exposure than does immediate feedback (Kulik & Kulik, 1988). But if other forms of instructional support are included, such as the addition of a reading passage, the effects of the double exposure are likely to be mitigated. Delayed and immediate feedback would then produce comparable results. Both the interference and frequency-of-feedback views appear to provide valid explanations of feedback-timing effects and both are supported by research (More, 1969; Newman, Williams, & Hiller, 1974; Peeck & Tillema, 1978; Suber & Anderson, 1975). The role of text as instructional support for questions (i.e., test-acquisition vs. text-with-questions effects), however, has not been adequately investigated.

The literature on feedback also leaves questions unanswered regarding the relationship of lesson questions to posttest questions. Often, posttest questions are identical in form and wording to the lesson questions, a rote recognition condition that substantially restricts the degree to which results can be generalized to typical learning situations. Bormuth, Manning, Carr and Pearson (1970) demonstrated experimentally how posttest questions could be adapted from instructional reading passages to measure comprehension learning by transforming and paraphrasing text (also see Anderson, 1972). Using their approach, the present study was designed to compare the effects on learning of three types of feedback strategies (KCR, AUC, and delayed) applied to five levels of lesson questions differing in degree of relatedness to posttest questions. An additional variable was whether feedback treatments were presented with associated text passages or with no text (i.e., test-acquisition. Subjects were low-ability high school students enrolled in a summer preparatory program in science. The following hypotheses were tested.

1. The provision of feedback vs. no feedback would improve learning across all feedback strategies and questioning levels.

2. AUC feedback would become increasingly effective relative to other feedback forms as question level (disparity between lesson and posttest questions) increased, due to providing additional processing opportunities and review of the information to be learned.
3. The provision of text would become more facilitative as question level increased due to furthering understanding of the material through the provision of additional descriptions and explanations.

Method

Subjects and Design

Subjects consisted of 100 eleventh grade students enrolled in a five-week CBI summer enrichment program sponsored by Memphis Partners Incorporated. Memphis Partners selects students considered to be at-risk from all schools in the metropolitan area. All subjects voluntarily participated in this study which was described as an American College Test (ACT) preparation course. All subjects were black; their median age was 17. To qualify for the program, they needed to (a) be entering the 12th grade in the fall; (b) have low ACT scores (between 10 and 15); and (c) be described by their guidance counselors and teachers in a written recommendation as having academic potential for college, despite their low standardized achievement scores.

Subjects were randomly assigned to one of 10 treatment groups consisting of five feedback conditions (KCR, AUC, delayed KCR, questions only, and no questions) crossed with two conditions of instructional support (text and no-text). Within-subjects factors consisted of five question levels (verbatim-identical, inferential-identical, inferential-transformed, inferential-paraphrased, and transformed-paraphrased), and two testings (immediate and retention). The analytical design thus consisted of a 5(feedback) x 2(instructional support) x 5(question level) x 2(testing) mixed analysis of variance (ANOVA).

Prior to the start of the instructional phase of the study, all subjects were administered the ACT Natural Science Reading Test and the Nelson-Denny Reading Comprehensive Test Form E to assess treatment group equivalence. The former test is a measure of science knowledge, and the latter is a measure of reading ability. Analyses of pretest scores, using a 5(feedback) x 2(support) ANOVA, indicated no significant differences between treatment groups on either measure.

Instructional and Assessment Materials

Text passages. The reading materials were adopted from the ACT National Sciences Reading Test, 8223c. They included four text passages entitled "Solids," "Genetics," "Compressed Gas," and "Trojan Asteroids." The average number of words per passage was 350. All passages were presented in print form to allow subjects continual access to them during the lesson and to create a more realistic learning situation. Readability of the passages, using Dale-Chall (1948) and Flesch (1948) procedures, ranged from 10th grade to college.

Lesson and posttest questions. For each passage, 10 lesson questions were constructed. Because the instructional orientation of the ACT passages and achievement test (ACT sample Test 8223C) emphasized inferential learning (i.e., reasoning from the passage to solve a problem or application), it was decided to make 8 out of each 10 (80%) lesson questions inferential and the remaining two questions (20%) verbatim. Inferential questions required going beyond the specific text information to formulate an idea or concept not explicitly stated. Verbatim questions repeated the text passage word for word.

Each of 40 lesson questions was made parallel to an existing verbatim or inferential posttest question adapted from the ACT Sample Test 8223c. Posttest questions were then varied in form on a random basis for the purpose of assessing different levels of learning. For the inferential questions a 2(transformation) x 2(paraphrase) design matrix was used to achieve these levels. As will be described below, one factor was whether a structural transformation or the original form of the corresponding lesson question was used; the other factor was whether paraphrased wording or original wording was used. Specifically, transformed posttest questions reversed the stem and the answer from the corresponding lesson question. To illustrate using a simple example, the question, "The capital of Arkansas is: (a) Little Rock (b) Memphis (c) Dallas" would be transformed to read: "Little Rock is the capital of (a) Arkansas (b) Tennessee (c) Texas." Paraphrased questions were constructed to maintain the same structure and meaning as corresponding lesson questions, but using different words or phrasing.

These manipulations resulted in five levels of posttest questions. Table 1 illustrates the five question forms in relation to a lesson text segment containing tested content. Both lesson questions and posttest questions were administered on a WICAT System 300 microcomputer with 30 student stations. The five forms are summarized below:

1. Verbatim-Identical (VI) tested verbatim learning using the same wording as the original text and the lesson questions.
2. Inferential-Identical (II) tested inferential learning using similar wording as the text and the identical wording as the lesson questions.
3. Inferential-Transformed (IT) tested inferential learning using similar wording as the text, but the question answer and stem were reversed relative to lesson questions.
4. Inferential-Paraphrased (IP) tested inferential learning using different words and phrasing relative to the text and the lesson questions.
5. Transformed-Paraphrased (TP) tested inferential learning using both a transformed structure and paraphrasing.

Insert Table 1 about here

Reliability of the posttest and lesson questions was assessed employing 94 high school students (47 for each question set). Split-half reliability, using the Spearman-Brown formula, was .66 for the posttest set and .85 for the lesson set.

Text and Feedback Treatments

Manipulation of the text variable involved either providing or not providing the passages to read during the lesson. The feedback variable consisted of three feedback forms and two no-feedback control conditions (questions without feedback and no questions). When presented without text, the no-questions variation represented a pure control condition in which students were administered the two posttests without receiving prior instruction. Each of the three feedback conditions is described below.

1. Knowledge of correct response (KCR). This condition, which was patterned after that used by Dempsey (1988; also Tait, Hartley, & Anderson, 1973), informed the learner of the correct answer after each response. Specifically, following a correct response, the word "RIGHT" was displayed at the bottom of the computer screen. Following an incorrect response, the word "WRONG" was displayed with the correct answer designated by an arrow. The learner was instructed to type the letter of the correct answer to continue.

2. Answer until correct (AUC). AUC, based on Dempsey (1988), provided the same feedback as KCR following correct responses. However, following the first incorrect answer to a given question, the prompt, "NO TRY AGAIN " was displayed at the bottom of the screen. The learner then made a second try, which if correct was followed by the usual "RIGHT," and if incorrect was followed by "WRONG" along with the instruction to type in the letter of the correct response (as designated by the arrow). Thus AUC was identical to KCR, except for the second try given following an initial error response.

Delayed feedback. This condition provided KCR-type feedback at the conclusion of all four lesson sections by individually presenting the 40 questions in original order, with the correct answer for each designated by an arrow. Separate from this concluding feedback display, an additional design consideration was whether to provide any immediate feedback to indicate the accuracy of responses. Given the difficulty and technical nature of the subject matter, we reasoned that the absence of such information would be frustrating to learners and unrealistic relative to what would probably be done by most designers in practice. Accordingly, we decided on a "middle ground" approach in which immediate feedback was provided, but the message was downgraded in "load" (information density) to KOR (as opposed to KCR); i.e., simply indicating that the answer was "RIGHT" or "WRONG," without designating the correct answer. Typically, the time delay from the learner's first response to the item to the delayed KCR was about 30 minutes.

Procedure

The summer preparation program continued for five weeks during students' school vacation. On selected weeks students attended experimental sessions, referred to as an "ACT prep course," for one hour at a convenient time during the day or night. Prior to their participation in the experiment, they had been administered the ACT Natural Science Reading Test to determine group equivalence. During Week 1 of the research period, they were administered the Nelson-Denny Reading Comprehension Test, as an additional measure of equivalence.

During Week 2, subjects participated in the treatment phase, receiving instruction appropriate to their assigned condition. Lesson questions were administered in blocks of 10 by computer in all conditions except the no-questions (posttest-only) treatment. Supporting text, where prescribed, was available at the learning station in print booklet form. The questions-and-text and no-text treatments were conducted on alternate days to avoid subjects becoming aware of the alternative experimental condition. Students in the questions-and-text condition were allowed to use the text in any way they wished. They were not given additional instructions relative to the no-text condition except for one sentence at the beginning of each block of questions, indicating that they should read a particular section "to help answer the questions." They were then left on their own to read and reference the text whenever and for as much time as they wanted. Observation of subjects reflected use of a variety of strategies, including reading the text first and then answering the lesson questions, reading the text as questions were answered, and/or referencing parts of the text following the completion of different questions. No text material was provided during the posttest or delayed posttest.

After subjects completed the assigned treatment, they were given a 10-minute break followed by the administration of the 40-item posttest. Subjects could spend as much time as they needed to complete the instructional phase; most finished in 45 minutes to 1 hour. Two weeks later, they were readministered the posttest unannounced to assess retention. The text portions were not available during either testing.

Results

The analyses of achievement scores used a 5 x 2 x 2 mixed ANOVA on each question level. Between-subjects factors were five feedback conditions (KCR, AUC, delayed, no-feedback, no-questions) and two instructional support conditions (text vs. no-text). The within-subjects factors were five question levels (VI, II, IT, IP, and TP) and two testings (posttest and retention test). Means for all conditions are shown in Table 2.

Insert Table 2 about here

The ANOVA yielded significant main effects due to feedback, F (4, 90) = 13.96, p, < .001, MSe = 5.99; and question level, F (4, 360) = 135.32, p < .001, MSe = .92. Each of these effects was qualified by significant interactions. Two-way interactions that reached significance were feedback x question, F (16, 360) = 13.50, p, < .001, MSe = .92; support x question, F (4, 360) = 2.46, p < .05, MSe = .92; and question x testing, F(4, 360) = 28.60, p < .001, MSe = 1.19. Significant three-way interactions were feedback x question x testing, F (16, 360) = 4.06, p < .001, MSe = 1.19; and support x question x testing, F(4, 360) = 5.20, p < .001; MSe = 1.19. Further, the four-way interaction also reached significance, F (16, 370) = 2.29, p < .003, MSe = 1.19.

Interpretation of the latter interaction, which qualifies all other effects, is obviously complicated by the four factors and 100 means it encompasses. Given that every interaction involving questions x testing was significant, it seemed appropriate for simplifying the interpretation of interaction patterns to conduct, as follow-up analyses, separate feedback x support x question ANOVAs for each test (immediate and retention). Due to the large number of factors involved in these analyses, the .01 level was used in judging significance. Results from each analysis are summarized in Table 3 and reported in the sections below.

Insert Table 3 about here

Immediate Test

As shown in Table 3, the feedback ($p < .001$) and question level ($p < .001$) main effects, but not the support main effect, were significant. Significant interactions were feedback x question level ($p < .001$), and support by question level ($p < .01$).

The feedback main effect was further analyzed via Tukey follow-up comparisons of the five overall treatment means. Results indicated that all three feedback strategies, KCR (M=4.7), AUC (M = 4.7), and delayed (M=4.5), were superior ($p < .05$) to both the no-feedback (M=3.2) and no-questions (M=3.0) control strategies. Similarly, Tukey follow-up tests of the question level main effect showed means on the two identical question forms, verbatim-identical (M=4.1) and inferential-identical (M=5.2), to surpass ($p < .05$) the means on the three reworded question levels, inferential-paraphrased (M=3.5), inferential-transformed (M=3.4), and transformed-paraphrased (M=3.1).

The significant feedback x question interaction ($p < .01$) reflected a general pattern for larger differences favoring feedback means over control means to occur on identical questions (VI and II) than on reworded questions (see Figure 1). Followup analyses involved comparing the five feedback means, using a Tukey test, for each type of question. The .01 level of significance was used to reduce the overall Type I error rate. Findings indicated that on both verbatim-identical and inferential-identical questions, each of the three feedback groups (KCR, AUC, and delayed) significantly surpassed each of the control groups (no-feedback and no-questions). On inferential-paraphrased questions, the only significant difference was that AUC surpassed no-questions. No differences were found on either inferential-transformed or transformed-paraphrased questions, although on the former measure, the differences favoring the highest group, AUC, over both of the control groups, approached significance ($.01 < p$'s $< .05$).

Insert Figure 1 about here

Followup analyses of the support x question level interaction ($p < .01$) consisted of comparing the text-and-question mean to the questions-only mean on each of the five question levels (see immediate test column means on Table 2). Multiple t tests, each using a .01 significance level, were used. Findings revealed that the only significant effect occurred on verbatim-identical questions, with text-and-questions (M=5.5) surpassing questions-only (M=4.6).

Retention Test

The same main effects and interactions that were significant on the immediate test were also significant on the retention test, with the exception of the support x questions interaction ($p > .01$; see Table 2). The feedback main effect ($p < .001$) was further analyzed by Tukey tests. As occurred for the immediate test, KCR (M=3.9), AUC (M=4.0), and delayed (M=38) feedback each surpassed no-feedback (M=3.1) and no-questions (M=3.0). Followup analyses of the question-level main effect ($p < .001$) indicated that scores on verbatim-identical (M=4.2) and inferential-identical (M=4.3) questions were higher than those on transformed-paraphrased (M=2.9), inferential-transformed (M=3.2), and inferential-paraphrased (M=3.3) questions. The transformed-paraphrased mean was significantly lower than each of the other question-level means.

The significant feedback x questions interaction ($p < .001$) was further analyzed by comparing the five feedback means, using a Tukey test, for each type of question (alpha = .01). The interaction is graphically displayed in Figure 2. Findings indicated that on verbatim-identical questions, the two highest groups, delayed and KCR, surpassed the two lowest groups, no-questions and no-feedback; AUC did not differ from any other groups. On inferential-identical questions, all three feedback groups surpassed both control groups. On inferential-transformed questions, no significant differences occurred; the largest difference, that favoring AUC over no-questions, approached significance at the .01 level ($p < .05$). On inferential-paraphrased questions, the only significant difference was that AUC surpassed no-questions. On transformed-paraphrased questions, no differences occurred. The overall pattern revealed from these comparisons is similar to but not as strong as that for the immediate test, showing (a) larger differences favoring the feedback conditions over the control conditions on the two identical question types (VI and II) than on the three reworded question types (IT, IP, and TP), and (b) a tendency for AUC effects to be more positive relative to the other feedback treatments on reworded than on identical question types.

Insert Figure 2 about here

Completion Time

Lesson completion times were analyzed for subjects in the three feedback conditions using a 3(feedback) x 2(text) ANOVA. The text main effect, F (1,52) = 38.9, $p < .001$ was significant, confirming the expected longer completion times for subjects who received text (M = 57.4 min.) than for those who did not (28.9 min.). The feedback main effect approached significance, F (2,54) = 2.89, $p < .06$, with the ordering of means from highest to lowest being delayed (M = 50.5), AUC (M = 41.6), and KCR (M = 37.4).

Discussion

The results of this study supported the hypothesized benefits for learning of providing response feedback on embedded lesson questions. To most educators, this result would hardly be viewed as surprising. The effectiveness of feedback is a basic tenet of instructional theory that has been demonstrated countless times by researchers beginning with the classic verbal learning studies by Thorndike (1931) on the effects of saying "Right" or "Wrong" following a subject's response. Frequent and consistent use of feedback is also strongly promoted in today's textbooks on teaching and educational psychology (e.g., Woolfolk, 1990, pp. 543-545; Slavin, 1988, pp. 383-387). But, while the benefits of feedback in general might be taken for granted, uncertainty still exists regarding how to select and optimize uses of different forms of feedback depending on characteristics of students and the learning situation.

As suggested from the present results, one important factor influencing feedback effects is the type of questioning employed, a variable that has typically not been controlled in previous studies. Had only one level of questioning been used, our findings would have been directly dependent on the particular level selected. Given the broader perspective obtained by manipulating five questioning levels, we were able to detect several basic trends. One was for feedback benefits to decrease as the similarity of posttest questions to corresponding lesson questions decreased. In other words, larger feedback effects occurred on the "identical" items than on the reworded ones. Another, in support of Hypothesis 2, was for the relative benefits of AUC feedback to increase as posttest question similarity decreased. Third, feedback effects relative to the control conditions tended to be greater without text than with text.

Better understanding of these outcomes can be obtained by analyzing the nature of the instructional support provided by the different feedback conditions. As suggested here and in previous studies involving identical lesson and posttest questions (Kulhavy 1977; Kulhavy & Anderson 1972; Smith 1988), the most direct benefit of KCR-type feedback (whether immediate or delayed) is informing learners of the correct answers to lesson questions. Thus, even when the level of learning does not extend beyond rote memorization, the benefit should be an increased ability to reconstruct those associations and identify the answers when the same questions appear again on a lesson posttest. Looking again at Figure 1, that effect is reflected by the three feedback groups' greater superiority over two control groups in the two conditions where posttest questions were exact replications of lesson questions (verbatim-identical and inferential-identical).

It has further been proposed that for strengthening associations between questions and correct answers, delayed feedback is especially advantageous by providing a second exposure to the item presentation at the end of the lesson (Kulik & Kulik, 1988) and by reducing proactive interference (Kulhavy & Anderson, 1972). Consistent with the emphasis of these explanations on rote-learning processes (i.e., connecting specific answers to associated questions), delayed-feedback effects have primarily been found in situations involving identical lesson and criterion test items (e.g., Kulik & Kulik, 1988; Sturgis, 1978; Suber & Anderson, 1975). Similarly, on both the immediate and retention tests in the present study, delayed-feedback was directionally higher than all other feedback conditions on verbatim-identical questions, and was significantly higher than the control conditions on verbatim-identical and inferential-identical questions only. Looking at Figures 1 and 2, the effectiveness of delayed feedback relative to the other conditions tended to decline as a general pattern as lesson-posttest question similarity decreased.

Despite its informational properties, feedback by itself does not necessarily increase depth-of-processing of the material being learned. In fact, a possible disadvantage of feedback may be that of supplanting the natural tendency of questions to stimulate information processing or "mathemagenic activity" (Rothkopf, 1966), as the learner searches memory or the text to find the answers to questions. That is, once the correct answer is identified, the learner may resort to memorizing or merely acknowledging it without engaging in further processing. In a similar vein, Andre (1979) discussed how the availability of feedback in text can short circuit the instructional effects of adjunct questions by allowing subjects to pe ek ahead at the answers and thus avoid searching the text to find the m on their own. Relevant to these interpretations, feedback effects were noticeably smaller for the reworded question forms than for the identical forms. It thus appears that feedback, especially KCR and delayed, generally did not stimulate deeper processing of the present material, while promoting only a limited degree of transfer to questions testing the same information as the lesson questions but differing in phrasing or structure.

From an information processing perspective, AUC feedback would appear to offer potential advantages over KCR as a result of requiring continued involvement with a question following an incorrect response (Dempsey & Driscoll, 1989; Noonan 1984). Such activity can increase depth-of-processing for the item (Smith, 1988), provided that the learner is not just guessing randomly (Underwood, 1963). On the present task, AUC tended to be effective relative to both KCR and delayed feedback on reworded questions, but was relatively ineffective on identical questions (see Table 2). This pattern suggests that AUC may have served to promote higher-order learning of the material, as learners reconsidered the questions they missed in light of their previous error responses and the remaining alternatives. Further research is needed to explore this possible function of AUC as well as to reconcile the mixed findings regarding AUC effects reported in previous studies (cf, Angell, 1949; Clariana, 1990; Dempsey & Driscoll, 1989; More, 1969).

That feedback effects tended to be stronger in the no-text than in the text condition seems predictable, given that the former subjects were completely dependent on the adjunct questions to learn the information. This result should not be interpreted to imply that learning is as good or better from questions only without accompanying text. Although there is little doubt that tests can teach (e.g., Fisher, Williams, & Roth, 1981; Meyer, 1965; Pressey, 1926, 1950), what is learned will be restricted by the particular focus of the items that happen to be included. The implication of this idea, along with our earlier discussion of feedback effects, is that feedback studies that employ identical lesson and posttest questions narrow the content domain to those specific questions, and, in the process, maximize the importance of the questions (and accompanying feedback) while minimizing the value of contextual support (e.g., text).

Another aspect of the present text versus no-text comparison was the failure to support the predicted tendency (Hypothesis 3) for text to become more facilitative as question level increased. In fact, the significant support x question interaction obtained on both tests reflected the opposite pattern; for example, the largest difference favoring text over no-text occurred on the lowest level question type, verbatim-identical. The suggestion is that subjects' processing of the text might have been at a fairly low level. An important factor in this regard appears to be the difficulty of the material and high reading level of the passages.

In summary, the major findings evidenced in the present research were that: (a) feedback was generally effective for learning, but more so on the lower-level (identical) questions than on the higher-level (reworded) ones; (b) feedback information had greater impact in the absence of supporting text than with supporting text; (c) relative to other treatments, AUC feedback tended to increase in effectiveness and delayed feedback to decrease in effectiveness as question level was varied from identical types to reworded types.

Past feedback studies, including the present investigation, have focused primarily on comparing learner achievement under different feedback strategies. Follow-up research that gives greater focus to intervening learning behaviors (e.g., degree of task engagement, referencing of text, note-taking) would shed light on the question of how information processing and study activity are influenced by those strategies. The present completion time results, for example, are suggestive of varied degrees of task engagement that occurred in the three feedback conditions. Acquiring better understanding of such processes should help to identify ways of using feedback more effectively to increase the range and degree of learning from embedded lesson questions.

References

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42(2), 145-170.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 49(2), 280-318.

Angell, G.W. (1949). The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry. Journal of Educational Research, 42(5), 391-394.

Bormuth, J.R., Manning, J., Carr, J., & Pearson, D. (1970). Children's comprehension of between- and within-sentence syntactic structure. Journal of Educational Research, 42(5), 391-394.

Clariana, R.B. (1990). A comparison of AUC feedback and KCR feedback under two conditions of contextualization. Journal of Computer-Based Instruction, 17(4), 125-129.

Dale, E., & Chall, J. (1948, January). A formula for predicting readability. Educational Research Bulletin, 27, pp. 11-20, 37-54.

Dempsey, J.V. (1988). The effects of four methods of immediate corrective feedback on retention, discrimination error, and feedback study time in computer-based instruction. Dissertation Abstracts International, 49(6), 1434.

Dempsey, J.V., & Driscoll, M.P. (1989). The effects of four methods of immediate corrective feedback on retention, discrimination error, and feedback study time in computer-based instruction. A paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Dick, W., & Latta, R. (1970). Comparative effects of ability and presentation mode in computer-assisted instruction and programmed instruction. Audio-Visual Communication Review, 18(3), 34-45.

Fisher, K., Williams, S., & Roth, J. (1981). Qualitative and quantitative differences in learning associated with multiple-choice testing. Journal of Research in Science Teaching, 18(5), 449-464.

Flesch, R.F. (1948). A new readability yardstick. Journal of Applied Psychology, 32, 221-233.

Gilman, D.A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. Journal of Educational Psychology, 60(6), 503-508.

Guthrie, E.R. (1935). The psychology of learning (1st ed.) New York: Harper & Brothers.

Hannafin, M.J., & Reiber, L.P. (1989). Psychological foundations of instructional design for emerging computer-based instructional technologies: Part I. Educational Technology Research and Development, 37(2), 91-101.

Kulhavy, R.W. (1977). Feedback in Written Instruction. Review of Educational Research, 47,(1), 211-232.

Kulhavy, R.W., & Anderson, R.C. (1972). Delay-retention effect with multiple-choice tests. Journal of Educational Psychology, 63(5), 505-512.

Kulhavy, R.W., & Stock, W.A. (1989). Feedback in written instruction: The place of response certitude. Educational Psychology Review, 1, 279-308.

Kulik, J.A., & Kulik, C. (1988). Timing of feedback and verbal learning. Review of Educational Research, 58(1), 79-97.

Merrill, J. (1985). Levels of questioning and forms of feedback: Instructional factors in courseware design. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago 1985. (ERIC Document Reproduction Service No. ED 266 766)

Merrill, J. (1987). Levels of questioning and forms of feedback: Instructional factors in courseware design. Journal of Computer-Based Instruction, 14, 18-22.

Meyer, D.E. (1965). Adjunct to self-study for aircrew refresher training under operational conditions in the Air Defense Command. Aerospace Medical Research Labs, Wright-Patterson AFB, OH. (ERIC Document Reproduction Service No. ED 131 830)

More, A.J. (1969). Delay of feedback and the acquisition and retention of verbal materials in the classroom. Journal of Educational Psychology, 60(5), 339-342.

Newman, M. I., Williams, R.G., & Hiller, J.H. (1974). Delay of information feedback in an applied setting: Effects on initially learned and unlearned items. Journal of Experimental Education, 42(4), 55-59.

Noonan, J.V. (1984). Feedback procedures in computer-assisted instruction: Knowledge-of-results, knowledge-of-correct-response, process explanations, and second attempts after explanations. Dissertation Abstracts International, 45(1), 131.

Peeck J., & Tillema, H.H. (1978). Delay of feedback and retention of correct and incorrect responses. Journal of Experimental Education, 38, 139-145.

Pressey, S.L. (1926). A simple apparatus which gives tests and scores - and teaches. School and Society, 23, 373-376.

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

Rothkopf, E.Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 3(4), 241-249.

Schimmel, B.J. (1983). A meta-analysis of feedback to learners in computerized and programmed instruction. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada. (ERIC Document Reproduction No. ED 233 708)

Schimmel, B.J. (1986). Feedback use by low-ability students in computer-based education. Dissertation Abstracts International, 47,(11), 4068.

Slavin, R.E. (1988). Educational psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Smith, P.L. (1988). Toward a taxonomy of feedback: Content and scheduling. A paper presented at the Annual Meeting of the Association for Educational Communications and Technology, New Orleans, LA.

Spock, P.A. (1987). Feedback and confidence of response for a rule-learning task using computer-assisted instruction. Dissertation Abstracts International, 48,(5), 1109.

Sturges, P.T. (1978). Delay of informative feedback in computer-assisted testing. Journal of Educational Psychology, 70(3), 378-387.

Suber, J.R., & Anderson, R.C. (1975). Delay-retention effect in natural classroom settings. Journal of Educational Psychology, 67(2), 170-173.

Tait, K., Hartley, J.R., & Anderson, R.C. (1973). Feedback procedures in computer-assisted arithmetic instruction. British Journal of Educational Psychology, 43, 161-171.

Thorndike, E.L. (1931). Human learning. New York: Appleton-Century, Crofts.

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, 65(3), 167-173.

Underwood, B.J. (1963). Stimulus selection in verbal learning. In C.J. Cofer and B.S. Musgrave (Eds.), Verbal behavior and verbal learning: problems and processes. New York: McGraw-Hill.

Waldrop, P.B., Justin, J.E., & Adams, T.M. (1986). A comparison of three types of feedback in a computer-assisted instruction task. Educational Technology, 26, 43-45.

Woolfolk, A. (1990). Educational psychology for teachers. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Table 1

Text Portion and Corresponding Questioning Levels

Text portion

In conductors, the electrons move easily; in insulators, they do not. Since moving electrons carry energy as well as charge, a good electrical conductor normally is also a good heat conductor, and an electrical insulator is also a poor heat conductor.

Questioning Levels

Verbatim Identical [a]

Normally, a good electrical conductor is also:

a. a good heat conductor.
b. a poor heat conductor.
c. a good electrical insulator.
d. the best electrical insulator.

Inferential Identical [b]

Copper is a poorer heat conductor than silver, then copper probably:

a. has a smaller heat capacity than silver.
b. is a poorer electrical conductor than silver.
c. is a semiconductor.
d. is a better electrical conductor than silver.

Inferential Transformed [c]

Copper is a poorer electrical conductor than silver, then copper probably:

a. is a semiconductor.
b. has a smaller heat capacity than silver.
c. is a poorer heat conductor than silver.
d. is a better heat conductor than silver.

Inferential Paraphrased [a]

Copper does not move heat energy as well as silver, so copper probably:

a. is a semiconductor.
b. moves electrical energy better than silver.
c. can hold less heat energy than silver.
d. does not move electrical energy as well as silver.

Inferential Transformed & Paraphrased [a]

Copper does not move electrical energy as well as silver, so copper probably:

a. moves heat energy better than silver.
b. is a semiconductor.
c. does not move heat energy as well as silver.
d. can hold less heat energy than silver.

Note: The letter of the correct answer is underscored in each item.
a Fabricated question used for illustrative purposes
b Actual lesson question for the text passage shown
c Actual posttest question for the text passage and lesson question shown

Table 2

Treatment Posttest and Retention Test Means of the Five Question Levels

Question Levels and Support

Feedback and Testing	Verbatim Identical		Inferential Identical		Inferential Transformed		Inferential Paraphrased		Transformed Paraphrased		Row Means Overall
	Q[a]	QT[b]	Q	QT	Q	QT	Q	QT	Q	QT	
Immediate Test											
KCR	5.9	6.1	6.5	7.1	4.6	3.2	2.7	4.1	3.4	3.6	4.7
AUC	5.5	5.[illegible]	6.7	5.5	4.4	3.6	4.8	3.7	4.0	3.2	4.7
Delayed	6.4	6.4	6.1	6.7	3.5	3.6	3.4	3.5	2.3	2.9	4.5
No-feedback	2.8	4.8	3.3	2.8	2.9	2.5	3.1	3.4	3.4	3.3	3.2
No-question	2.4	4.9	3.6	3.4	2.5	2.7	2.6	3.2	2.4	2.5	3.0
Column means	4.8	5.5	5.2	5.1	3.6	3.1	3.3	3.6	3.1	3.1.	4.0
Retention Test											
KCR	4.4	4.8	5.0	5.4	3.2	3.4	3.3	3.7	2.4	3.0	3.9
AUC	4.4	4.4	5.6	4.5	4.0	3.2	4.1	3.5	3.1	3.2	4.0
Delayed	5.2	5.0	4.9	5.2	3.4	3.3	3.3	2.8	2.7	2.3	3.8
No-feedback	2.8	4.1	3.1	3.2	2.9	2.8	2.9	3.0	3.2	3.2	3.1
No-question	2.6	4.4	2.9	3.2	2.4	2.9	2.7	2.8	2.8	2.9	3.0
Column Means	3.9	4.5	4.3	4.3	3.2	3.1	3.3	3.2	2.8	2.9	3.6

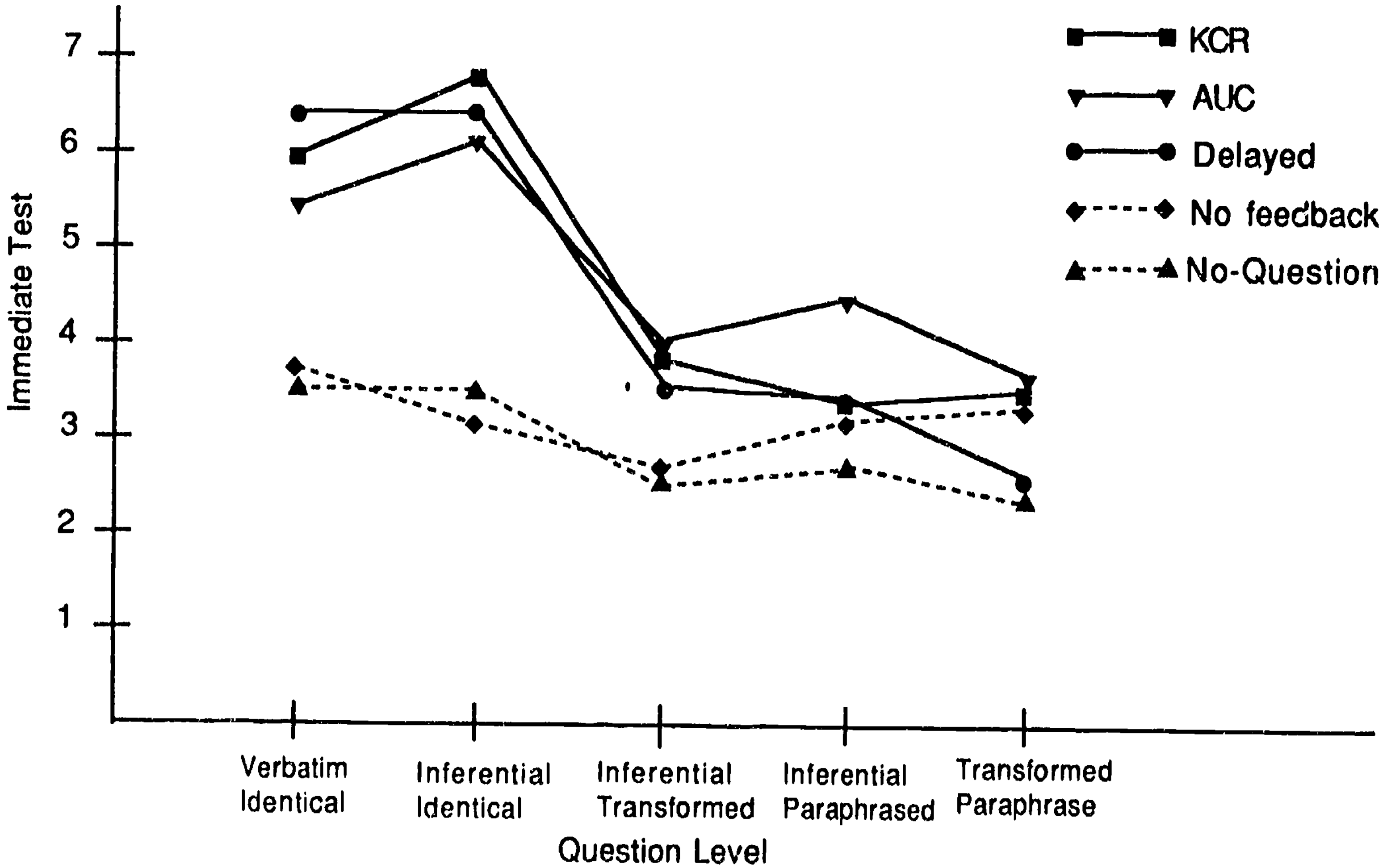
Note. Scores could range from 0-8 in each cell.
[a]Q = Questions only; [b]QT = Text and Questions

Table 3

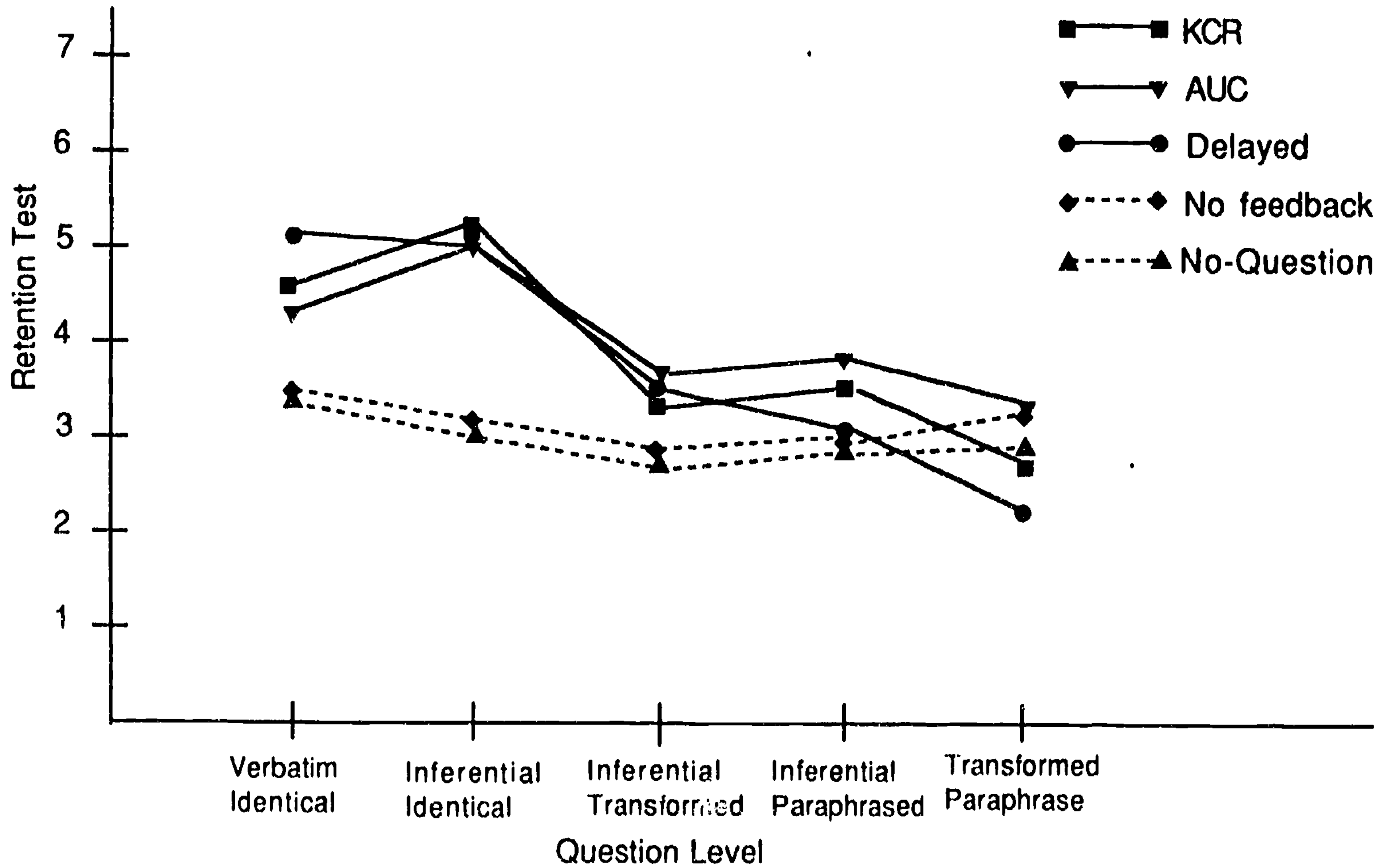
Results of Feedback x Support x Question Level ANOVAs by Testing

		Testing			
		Immediate		Retention	
Source of Variance	df	MS	F	MS	F
Between Subjects					
Feedback (F)	4	68.03	16.05**	22.18	9.13**
Support (5)	1	1.92	.45	1.80	.74
F x S	4	7.57	1.78	4.79	1.97
Error	90	4.24		2.43	
Within Subjects					
Question Level (Q)	4	101.27	68.78**	43.08	65.27**
F x Q	16	10.97	7.45**	5.00	7.58**
S x Q	4	6.43	4.37*	2.18	3.31
F x S x Q	16	2.65	1.80	.90	1.36
Error	360	1.47		.66	

* $p < .01$
** $p < .001$


7
6
5
4
3
2
1
Immediate Test
KCR
AUC
Delayed
No feedback
No-Question
Verbatim Identical
Inferential Identical
Inferential Transformed
Inferential Paraphrased
Transformed Paraphrase
Question Level

174



KCR
AUC
Delayed
No feedback
No-Question
Retention Test
7
6
5
4
3
2
1
Verbatim Identical
Inferential Identical
Inferential Transformed
Inferential Paraphrased
Transformed Paraphrase
Question Level