AUTHOR: Roy B. Clariana; Daren Wagner; Lucia C. Roher Murphy
TITLE: Applying a Connectionist Description of Feedback Timing
SOURCE: Educational Technology Research and Development v48 no3 p5-21 2000

ABSTRACT

This investigation considers the effects of feedback on memory with an emphasis on retention of initial error responses. Based on a connectionist model (Clariana, 1999a), this study hypothesized that delayed-retention memory of initial lesson responses would be greater for delayed feedback compared to immediate feedback, that feedback effects will be greatest with difficult items, and that there would be a disordinal interaction of feedback timing and item difficulty. High school students (n = 52) completed a computer-based lesson with either delayed feedback, single-try immediate feedback, or multiple-try immediate feedback. There was a significant difference for type of feedback, with retention test memory of initial lesson responses greater under delayed feedback than under immediate feedback. Also, instructional feedback effects varied depending on lesson item difficulty. The findings indicate that a connectionist model can explain instructional feedback effects.

* Learning involves the interaction of new information provided by instruction with existing information already in the learner's memory (Ausubel, 1968; Bruner, 1990). When a learner commits to a lesson response, that response reflects the learner's immediate understanding of that instructional instance. Lesson responses, and especially initial lesson responses (ILRs) are a useful and interesting measure of a learner's existing information. During learning, when an ILR is the correct response, feedback should confirm and strengthen that memory trace. When an ILR is the incorrect response, corrective feedback refutes the ILR. Current models of the effects of feedback focus on what happens to memory associations that correspond to the correct response, but there are few data and no generally accepted theory-based explanation for what effect feedback has on ILR errors. It is plausible that the memory trace of an initial error may be weakened because it is an error, it may remain unaffected, or less likely an initial error may be strengthened, perhaps because it has been brought to the attention of the learner.

Describing what happens to memory traces of ILR errors is necessary for determining whether ILR errors interfere with attaining correct responses, and so is a key to understanding how feedback works. During retention tests, learners can be asked to identify their ILRs as well as indicating the correct responses. Lesson ILR data can then be compared to the learners' retention test memory of their ILRs. This comparison would indicate whether ILR memory traces are weakened, strengthened, or remain unchanged as a result of the lesson feedback. These data can also be compared to memory of correct responses, thus indicating how ILRs, especially errors, may affect learning correct responses.

One focus of this investigation is on feedback timing especially as applied in computer-mediated learning. Feedback can be provided immediately after a learner's response, or can be delayed for some amount of time such as seconds, minutes, or days later. Many feedback timing studies and meta-analyses of studies suggest guidelines for when to use delayed and immediate feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kulik & Kulik, 1988). For example, in situations that use test items with feedback as instruction, delayed feedback is superior to immediate feedback with an effect size of 0.36, while in studies that use questions with feedback and include additional instructional materials, like text, immediate feedback is superior to delayed feedback with an effect size of 0.28 (Kulik & Kulik, 1988). These are general guidelines; often there are no significant differences in learning outcomes for immediate versus delayed feedback.

However, though feedback timing is a data-rich research area, the theoretical underpinnings are disputed and inconclusive. Kulhavy and Stock (1989) while considering existing feedback research have stated:

As with item feedback research, there has been no attempt to develop either a logic or a metric for creating instruction-based elaboration feedback. At this point, there are simply lots of studies testing whatever feedback configurations happen to be handy. (p. 288)

Systematically accounting for ILRs should provide a better understanding of the effect of feedback timing.

A second focus of this investigation is to apply a connectionist approach to explain feedback effects. Connectionist models have some semblance of biological plausibility and exhibit the crucial ability to learn by experience. For these and other reasons, connectionist approaches provide another way to think about human memory and learning. Shanks (1995) says, “Connectionism has had an enormous impact in the last decade across the whole field of human learning, from perceptual-motor learning to language acquisition” (p. 107). A fundamental connectionist learning rule, the delta rule, may provide a mathematical account of the effects of feedback on learning. If so, this investigation differs substantially from previous studies of feedback timing by using the delta rule for the first time to predict both direction and magnitude of learning from different forms of feedback.

Following below, we briefly review two explanations of feedback timing effects, the Interference Perseveration Hypothesis (IPH, Kulhavy & Anderson, 1972), which is based on inhibitory activity of ILRs, and a more general information-processing explanation that is based on a dual-trace assumption (Kulik & Kulik, 1988). Then, a connectionist description of feedback timing effects based on the delta rule is provided. This section concludes with the purpose of the study.

FEEDBACK TIMING

Immediate and delayed feedback differ in terms of the timing of feedback presentation as well as in the number of stimulus exposures provided. Typically, delayed feedback clearly presents at least two exposures with each item, the first during the learner’s initial response, and the second after some time delay when the feedback containing the stimulus and correct response are presented. Immediate feedback provides the response opportunity and the feedback within one instance. It is unclear whether feedback timing or number of explicit stimulus exposures most affects previously observed differences for immediate and delayed feedback.

It may be possible to separate the effects of immediacy versus multiple-exposures by including a type of feedback, multiple-try immediate feedback (MTF), which has aspects of both (Bangert-Drowns et al., 1991; Clariana, 1993). MTF is like the more common form of immediate feedback (single-try feedback, STF) in terms of immediacy of feedback timing. But in terms of multiple explicit stimulus exposures, multiple-try feedback is like delayed feedback (DF), at least with lesson errors. The present study includes MTF, STF, and DF in order to examine the effects of timing interval and multiple exposures on retention test memory of both ILRs and correct responses. It is expected that factors related to timing (such as interference) are of central importance in describing the difference between delayed and immediate feedback. If so, MTF retention test responses should be more like STF than DF. However, if factors related to number of stimulus exposures are central (such as multiple competing memory traces), then MTF will be more like DF.

Much attention has been given to feedback timing due in part to a phenomenon called the Delay Retention Effect (DRE). The DRE is an interesting and fairly reliable phenomenon that occurs with delayed feedback when multiple-choice questions with feedback are used as instruction. On identical item immediate posttests, immediate and delayed feedback groups obtain equal scores, but on retention test measures (usually 24 hours or a week later), the delayed feedback group surpasses the immediate feedback group (Brackbill, Blobitt, Davlin, & Wagner, 1963; Swindell & Walls, 1993). There are at least two plausible explanations of feedback timing and the DRE, the IPH (Kulhavy & Anderson, 1972) and a dual-trace information processing explanation (Glover, 1989; Kulik & Kulik, 1988).

IPH proposes that initial lesson errors interfere with learning the correct response through proactive inhibition. With immediate feedback, the learner strongly remembers the ILR error and so interference between the ILR error and the feedback-provided correct response is strong, inhibiting acquisition of the correct response. But with delayed feedback, the learner has somewhat forgotten the ILR error, possibly because of learning decay over time, and so interference between the ILR and the correct response is reduced allowing for better acquisition of the correct response. To summarize, the IPH describes an inhibitory process due to ILR errors that interferes with learning correct responses, but does not clearly explain why ILR errors may be forgotten, even though what happens with the ILR error memory trace is central and critical to the IPH.

Research supporting the IPH is mixed (Kulhavy & Anderson, 1972; Kulhavy & Stock, 1989; Markowotz & Renner, 1966; Rankin & Trepper, 1978; Schroth & Lund, 1993; Sturges, 1969, 1972, 1978; Sturges, Sarafino, & Donaldson, 1968; Suber & Anderson, 1975). In fact, in some investigations, lesson errors were remembered quite well in both the immediate and the delayed feedback conditions (Peeck & Tillema, 1978). This suggests two questions. Do learners remember ILR errors during feedback presentation and retention testing? And, if ILR errors are remembered, do remembered errors interfere with learning the correct response?

Though not fully explicated, the dual-trace information processing explanation for feedback timing is possibly best summed up by Kulik and Kulik (1988). They state:

In the immediate feedback condition, the trials may be almost fused. In the delayed feedback situation, the two trials are separate. Delayed feedback then may appear to be better than immediate feedback because two separate trails are better than one. (p. 94)

According to this approach, when the ILR is correct, the learner in the delayed feedback group has two opportunities to process the correct response; thus learners receiving delayed feedback will strongly remember their ILRs that are correct. When the ILR is incorrect, the learner in the delayed feedback condition will have a memory trace of the ILR error that results from making the response and another memory trace of the correct response that results from the feedback presentation. On a later multiple-choice retention test, the learner will remember the ILR error response (to some level) and will also remember the correct response (to some level), and the two interact at that moment. The learner may or may not realize that the ILR error is an error, however, a memory system that remembers errors to some degree should have a decided selective evolutionary advantage. If the ILR error memory trace is "tagged" as an error, this will increase the probability of selecting the correct response on a multiple-choice retention test. Alternately, if an ILR error is not remembered as an error, it is likely to reappear as a retention test error, a commonly occurring phenomenon termed error perseveration (Kulhavy & Anderson, 1972).

With some qualifications, both models can account for the superior DRE for delayed over immediate feedback on delayed retention tests. With IPH, interference occurs late during encoding (that is, sometime after the immediate posttest or else the advantage for delayed feedback would also occur on the immediate posttest), while the dual-trace explanation likely involves retrieval response competition between the dual memory traces. The trace with the greatest activation level should be selected as the presumed correct answer.

To summarize, the IPH depends on ILR errors interfering or not with learning the correct response (the error memory trace interferes with the encoding of the correct response memory trace), and the DRE then depends on ILR error memories being weaker and thus interfering less under delayed feedback compared to immediate feedback. Alternately, the dual trace explanation depends on multiple-processing opportunities with delayed feedback with resulting better memory both of the ILR error and of the correct response, increasing the probability of selecting correct responses on the retention test. Therefore, if a delayed feedback group can remember ILR errors better than an immediate feedback group, this would support the dual-trace explanation; while if the immediate feedback group remembers ILR errors better than the delayed feedback group, this would tend to support the IPH. Accounting for ILRs, and especially if ILR errors are strengthened, weakened, or remain unchanged, lends support to either the IPH or the dual-trace explanation of feedback timing effects.

CONNECTIONIST DESCRIPTION OF FEEDBACK TIMING EFFECTS

Clariana (1999a) has suggested that a connectionist model can be used to predict posttest memory activation levels of ILR errors and of correct responses for immediate and delayed feedback. Clarifying what happens to ILR errors and how it happens is critical for describing interference effects and the possibility of dual traces. This investigation applies the connectionist model of feedback timing described by Clariana (1999a). But first, what is a connectionist approach?

Connectionist models apply various mathematical rules within neural network computer simulations in an effort, among other things, to mimic and describe human memory associations and learning. The theory includes several families of models, such as simple feedforward networks, pattern associators, multilayer networks with backpropagation, competitive networks, and recurrent networks, which differ slightly in how the nodes of the network are interconnected, but which differ greatly in the kind of processing that they are able to do (see McLeod, Plunkett, & Rolls, 1998, for more detail). Neural networks have been shown to be capable of pattern matching, pattern completion, retrieval by content, recognition, prototype extraction, and classification to name a few (Haberlandt, 1997).

For example, Seidenberg and McClelland (1989) trained a computer neural network to read aloud all English monosyllabic words (about 3,000 words). After 250 training epochs, the model could correctly pronounce 97% of the 3,000 words in the training set. This neural network was able to accomplish this task without a local lexical store and, more importantly, without being given a set of rules. Elman (1993) trained a neural network with sentences rather than words, and was able to show that the network could satisfy long-distance grammatical dependencies (matching syntax). Plunkett and Marchman (1993; 1996) have modeled early lexical development which parallels that observed in children. Their neural network model that produces past tense forms of regular and irregular verbs has challenged the language acquisition orthodoxy that language learning depends on both innate prewiring of the system and on learning symbolic rules of the language.

As applied in the present investigation, Clariana (1999a) selected the simplest connectionist model that is able to describe feedback timing effects, a simple feedforward network (see Figure 1). A feedforward network consists of input nodes connected to multiple output nods (Shanks, 1995). A node is an ensemble of neurons that act in concert. The output node with the strongest association to the input node represents the response the neural network will give to that input pattern.

Among a number of connectionist learning rules, the delta rule (Shanks, 1995; Widrow & Hoff, 1960) is one of the simplest and most common that includes the effects of feedback on learning. The delta rule describes the change in association weight, termed Δw, between an input unit and an output unit at each learning trail, as $\Delta W_{io}=\alpha a_{in}\ (t_o\text{-}a_{out})$, where α is the learning rate parameter, a_{in} is the activation level of input units, t_o is the desired response (the t refers to "teacher," in this case t_o is item feedback), and a_{out} is the activation level of the output units (Shanks, 1995). In instructional terms, learning is an increase in association, that is, an increase in w_{io} between the stimulus (a_{in}) and the correct response (a_{out}), with a relative decrease in association, that is, a decrease in w_{io} for incorrect responses.

To apply the delta rule in this study, following Clariana (1999a), this investigation assumes that lesson average item difficulty values are reasonable estimates of the association weights of the correct responses. Item difficulty (p) is defined as the proportion (p_g) of individuals who answer an item (g) correctly (item difficulty notation convention from Crocker & Algina, 1986). For example an item difficulty of .20 indicates that 20% of the learners responded correctly to that item. Item difficulty values range from 0.00 to 1.00 with low values indicating difficult items and high values indicating easy items. Using lesson average item difficulty values as a group's estimate of the initial lesson item activation (a_{out}) seems reasonable in that lesson item difficulty is the actual averaged probability of selecting that alternative as the correct response during the lesson for that population of learners. To our knowledge, this is the first investigation to utilize lesson item difficulty values as a measure of input and output activation levels.

In the delta rule equation, feedback impacts learning in the term (t_o - a_{out}). Customarily, the values for t_o and a_{out} are constrained between 0 and 1 (McLeod, Plunkett, & Rolls, 1998). The value for t_o equals 1 if the activation level of the output unit matches the desired response (i.e., with correct responses) and t_o equals zero if the activation level of the output unit does not match the desired response (i.e., with incorrect responses). So with correct responses the association weight increases since (1 - a_{out}) is positive, while with incorrect responses the association weight decreases since (0 - a_{out}) is negative. For example, assume for a moment that a_{out} is 0.74; for a correct response the term (t_o - a_{out}) would equal (1 - 0.74) = 0.26, a positive value with a resulting increase in association weight between the input and output units. For an incorrect response, the term would equal (0 - 0.74) = -0.74, a negative value with a resulting large decrease in association weight between the input and output units.

In other words, when feedback is provided as part of the responding instance, correct responses are strengthened and incorrect responses weakened. The amount of increment or decrement is determined by the delta rule, $\Delta W_{io} = \alpha a_{in}\ (t_o - a_{out})$. For example (per Clariana, 1999a, where αa_{in} = 0.4), an easy item such as P_g = .8 (and so initial a_{out} = .8) will receive only a small increment. Specifically ΔW_{io} = 0.4 × (1 - .8) = 0.08; so the value of a_{out} after feedback will be .8 + 0.08 = .88. A difficult item such as P_g = .3, will receive a large increment. Specifically ΔW_{io} = 0.4 × (1 - .3) = 0.28; so the value of a_{out} after feedback will be .3 + 0.28 = .58. Therefore, if the delta rule applies to feedback, feedback will have its greatest effect on more difficult items, a result that has been reported by Sturges (1978) and suggested by Bangert-Drowns et al. (1991). In the same way, the delta rule can predict decreases in association weight. For example, suppose a response with an initial a_{out} of 0.5 is an error. Per the delta rule, the association weight between the input unit and the error output unit would decrease. Specifically $\Delta W_{io} = 0.4 \times (0 - .5) = -0.20$; so the value of a_{out} after feedback will be .5 - 0.20 = .0.30.

Thus, given lesson item difficulties (initial a_{out}), the delta rule would be able to predict posttest item difficulties (a_{out} after feedback). What values would the delta rule provide for ILRs for immediate and for delayed feedback? In the present investigation, first the delta rule would predict that for correct lesson responses, memory of ILRs and of correct responses would be strengthened in general for both immediate and delayed feedback, since $t_o = 1$ and so ($t_o - a_{out}$) is positive. Second, for lesson errors, the ILR association with the item stem would be weakened for immediate feedback since $t_o = 0$ and ($t_o - a_{out}$) is negative, but not for delayed feedback.

For delayed feedback, the connectionist model would predict that ILR errors would actually be strengthened. In associative learning in living systems, there is a small window of time while the specific input pattern is activated lasting probably less than 4 sec (Shanks, Pearson, & Dickerson, 1989) when those associations can be strengthened or weakened. Immediate feedback provides the necessary teacher feedback information within this time frame while delayed feedback does not. Specifically, with delayed feedback, since corrective feedback is not immediately provided, the learning rule association process would act as though the error response is correct, thus strengthening the association weight of the error. Thus this connectionist model clearly predicts that ILR errors are strengthened by delayed feedback. This phenomenon would account for error perseveration with delayed feedback, where errors committed early in learning are very likely to reappear later (Kulhavy & Anderson, 1972).

Based on the connectionist model of feedback timing described by Clariana (1999a), several hypotheses can be stated. As described above, retention test memory of ILRs will be considerably greater for delayed feedback than for immediate feedback at all item difficulty levels (see dashed lines in Figure 2). Next also as described above, both types of feedback obtain the greatest lesson-to-posttest gain with difficult lesson items. Last, retention test memory of correct responses will vary across the range of possible lesson item difficulty values for the delayed and immediate feedback forms, with immediate feedback slightly better than delayed feedback with more difficult lesson items, and delayed feedback slightly better than immediate feedback with easier lesson items (see the solid lines in Figure 2).

PURPOSE

This investigation considers the effects of feedback on memory with an emphasis on retention of initial error responses. The following three hypotheses are based on a connectionist model described by Clariana (1999a):

1. Retention test memory of ILRs will be greater for DF compared to STF and MTF across all item difficulties, but especially with difficult lesson items. The dual-trace explanation makes the same prediction as the connectionist model, except that it can only suggest direction of the effect, while the connectionist model predicts both direction and magnitude of the difference between immediate and delayed feedback. Alternately, the IPH will have problems if DF acts to increase memory of initial lesson errors, since that would result in increased proactive interference for delayed feedback relative to immediate feedback. The IPH then must predict the opposite, that proactive interference is greater with immediate feedback relative to delayed feedback and so memory of ILR errors will be less for DF compared to STF.

2. All forms of feedback will have their greatest effects with difficult items as suggested by Bangert-Drowns et al. (1991). The connectionist model predicts both the direction and magnitude of this difference. Neither the dual-trace explanation nor the

IPH considers possible item difficulty effects, though if such occurs, it would neither support nor refute these explanations.

3. There is an interaction of feedback timing and item difficulty, specifically, immediate feedback will be better for difficult items and delayed feedback will be better for easy items. Alternately, the IPH would predict that with difficult items, since incorrect lesson responses are being selected at a higher rate than correct responses for that item, interference with learning the correct answer would likely be greater with immediate feedback but less for delayed feedback. However with easy items, the learner is generally selecting correct responses during the lesson, so there is little interference and thus no difference between immediate and delayed feedback. Thus the IPH would predict that delayed feedback would be superior to immediate feedback for difficult lesson items but this difference would be absent for easy items, a position that is exactly opposite to that predicted by the connectionist model. The dual-trace explanation does not consider item difficulty effects.

This study differs substantially from previous studies of feedback timing through its use of the delta rule to predict both direction and magnitude of instructional effect. Note that the connectionist model provides a lower level or more fundamental description of the instructional effects of feedback than the two information processing models.

METHOD

SUBJECTS

The available sample for this study included students from three high school social studies classes (n = 87) from a small town in a northeastern state. The students were mostly sophomores, with a few juniors. A number of students chose not to participate, some were absent, a few forgot to return their signed consent forms, and three students' data were dropped because they were incomplete, yielding a final sample of 52 students. The final sample contained more females (71%) than males.

MATERIALS AND PROCEDURE

The computer-based lesson material consisted of eight reading passages and 36 five-alternative multiple-choice questions from the Nelson-Denny Reading Test, Form E (Brown, Bennett, & Hanna, 1981; with permission of the publisher). These text passages and questions were chosen based on the quality of the text and questions, their high reliability, the availability of extensive test reliability metrics, and content appropriateness for this audience. Field trials by the developers involved more than 14,000 students (Brown et al., 1981). Currently, these materials are extensively used in the field and are readily available from the publishers.

The content of the eight reading passages was: the poetess Elizabeth Barrett Browning (637 words, 8.4 Flesch-Kincaid Grade level); the importance of marketing on the economy (193 words, 13.4 grade level); atomic energy (207 words, 11.8 grade level); public response to Mars exploration (212) words, 9.1 grade level); the importance of fungi (200 words, 11.7 grade level); local government (199 words, 7.7 grade level); the painters El Greco and Caravaggio (253 words, 13.1 grade level); and random assignment in experimental design (193 words, 14.6 grade level). In total, there were about 2,000 words of text and 36 questions, or about two questions per paragraph of text.

The 36 questions consisted of 18 verbatim questions and 18 inferential questions, similar to Peeck and Tillema (1978). Verbatim questions relate directly to one proposition in the instructional text. For example, the text may say, "Daytimes Robert Browning walked with Elizabeth's little dog Flush, but he seldom could be lured from his home evenings," with the associated verbatim question, "Flush was the name of the

Brownings': (a) cat, (b) dog,(c) canary, (d) gold fish, (e) thrush?" Inferential questions relate to multiple propositions in the text and can be answered by considering the passage as a whole. For example, the inferential question, "The authors of this passage placed most attention on the Brownings: (a) literary efforts, (b) personal relationship, (c) social contacts, (d) problems, (e) early meeting?" can be answered by considering the entire passage. Thus, inferential questions have many indirect connections to the instructional text, while verbatim questions have one or only a few direct connections to the text.

The computer-based lessons were developed in HyperCard and delivered on MacIntosh computers. The three alternative computer-based lessons were identical except for the type of feedback that was presented after the learner responded to multiple-choice questions. For all treatments, each of the eight text passages was presented in turn in scrolling text windows. After reading a text passage, the learner would respond to either four or eight five-alternative multiple-choice questions (the longer passages had more questions). The learner would then proceed to the next text passage reading the text and answering the questions at his or her own pace. Simple navigation buttons along the bottom of every screen allowed the learner to easily move back and forth between text and questions at any time.

The three alternate treatments were delayed feedback (DF), single-try immediate feedback (STF), and multiple-try immediate feedback (MTF). Note that since retention test memory of initial lesson responses is a dependent variable in this study, in order to prevent rehearsal of initial lesson responses, the final feedback screen for each item in all three treatments did not include item distractors (Sassenrath & Yonge, 1969; Sturges, 1969).

The STF treatment provided the correct response immediately after one learner response, whether the response was right or wrong. After a response, an arrow would point to the correct alternative and the learner was told "Right" when correct and "No, here is the answer" when wrong. In either case, the stem and correct answer were displayed, the item distractors were not shown. Like STF, the MTF treatment provided the correct response immediately after a correct response. However, with an incorrect response, the learner was told "No, try again" and continued to select answers until the correct response was selected, then an arrow would point to the correct response and the learner was told "Right." Then the stem and correct answer were displayed; the item distractors were not shown. Note that STF and MTF are identical when the learner's initial response is correct, but obviously differ when the learner's response is incorrect. The DF treatment required the learner to respond to questions and move on without any immediate feedback. After all text passage and questions were completed, then all of the items were presented again in the original order. Only the correct responses were shown with each question, the distractors were not shown, and the student could only read the display. Thus the DF feedback screen displays were identical to the final feedback screen displays provided for STF and MTF, only the DF feedback screen displays were presented after the entire lesson rather than immediately with each item.

The purpose and requirements of the study were explained in three classes of students all taught by the same teacher. Those students choosing to participate collected consent forms to be signed by a parent or guardian. About a week later, participants moved to the school computer room during their social studies class, and were randomly assigned to one of the three computer-based treatments, STF, MTF, or DF. One day later, participants completed the paper-and-pencil 24-hour retention test in class.

CRITERION MEASURES

The retention test given a day after the lesson was designed to measure memory of initial lesson responses and of the correct responses. This paper-based retention test used the same 36 multiple-choice items that were used in the computer-based lesson. These 36-items fell into two groups of 18 items each, verbatim and inferential. These two groups were further blocked into three categories by lesson item difficulty, easy (M = .87), midrange (M = .72), and difficult (M = .50), with each block containing six items. The associated reading passages were not made available to the students during retention testing. There were four or five questions on each 8 × 11″ page. Each question had two blanks, one labeled “1st” for the initial lesson response and one labeled “C” for the correct response.

The retention test contained the following instructions, “Note that each question has two blanks. The first blank is a check to see if you can remember the first answer that you gave during the computer lesson. The second blank is for the correct response to the question.” These instructions were read aloud by the teacher. The teacher answered questions about how to complete the test, and then students were given as much time as needed to finish.

The total amount of time spent completing the lesson was also of interest. Lesson time data for each student were collected by the computer program, and included total time from the display of the first screen until the student exited the lesson.

RESULTS

Dependent variables were lesson time, and retention test recognition memory of initial lesson responses and of correct responses with two kinds of questions (verbatim and inferential) across three levels of lesson item difficulty (difficult, midrange, and easy). These data were analyzed by separate analysis of variance (ANOVA), and probabilities were evaluated more conservatively using Greenhouse-Geisser and Huynh-Feldt corrections automatically provided by the SYSTAT 8.0 (1998) analysis package. Lesson time data and retention test means and standard deviations for each treatment group at each item kind and difficulty level are provided in Table 1.

Lesson scores were analyzed as a check of initial group equivalence. The random assignment was judged successful with overall lesson means of 24.4 (of 36 maximum) for the STF group, 25.4 for the MTF group, and 25.5 for the DF group. The comparison of these means using ANOVA was nonsignificant, $F(1, 51) = 0.43$, $p = .65$.

LESSON TIME

The means for total time spent in each lesson treatment were DF (1,899 seconds), MTF (1,554 seconds), and STF (1,577 seconds). The ANOVA on time data yielded $F(2, 49) = 5.39$, $p < .01$. Scheffé follow-up tests indicated that DF time was significantly ($p < .01$) greater than both MTF and STF times. These findings parallel those of previous studies (see Clariana, Ross, & Morrison, 1991) that learners take more time with delayed feedback than with immediate feedback, probably because of rereading and thinking about both the question stem and the correct response during DF feedback presentation, but only reading the correct response but not rereading the question stem during STF and MTF feedback presentations.

Utilizing lesson time as a covariate in analysis (ANCOVA) can statistically control for possible confounding effects of time. However, the simple correlation between lesson time and retention test variables was quite low (all $r < .20$), which precludes ANCOVA due to increase in Type II errors (Kennedy & Bush, 1985). In addition, ANCOVA requires that the covariable not be affected by the administration of the treatments (covariable-treatment independence), which is clearly not the case in this investigation. Evans and Anastasio (1968) have concluded that in such cases, treatment means may be over- or under-adjusted, with spurious results. Thus analysis using lesson time as a covariate was not utilized.

RETENTION TEST

Retention test data were analyzed by a mixed 3 × (2 × 2 × 3) ANOVA with one between-subjects factor, feedback condition (DF, MTF, or STF), and three within-subjects factors, (a) type of response (retention of initial lesson response and of the correct response), (b) kind of question (verbatim and inferential), and (c) lesson item difficulty block (difficult, midvalues, and easy). The interaction of feedback and type of response was significant, $F(2, 49) = 7.15$, $MSE = 0.023$, $p < .01$ (see Figure 3). The follow-up Scheffé test showed that the DF treatment group mean for ILR ($M = 0.84$) was significantly larger than the STF group mean for ILR ($M = 0.73$), no other mean comparisons were significant.

A significant effect was obtained for kind of question, $F(1, 49) = 8.29$, $MSE = 0.035$, $P < .01$, indicating that the verbatim retention test mean ($M = 0.80$) was greater than the inferential retention test mean ($M = 0.76$), which simply reflects the lesson values for verbatim ($M = 0.74$) and inferential ($M = 0.65$) questions. Next, a significant effect was obtained for question difficulty, $F(2, 98) = 32.20$, $MSE = 0.035$, $p < .01$. The retention test means for each item difficulty level are: easy items ($M = 0.86$), midrange difficulty items ($M = 0.77$), and difficult items ($M = 0.71$). As with kind of question above, these retention test values simply reflect lesson values, which are: easy items ($M = 0.87$), midrange difficulty items ($M = 0.72$), and difficult items ($M = 0.50$). Though these findings are significant, they have little practical meaning.

The three-way interaction of feedback, type of response, and question difficulty was significant, $F(4, 98) = 3.52$, $MSE = 0.015$, $p < .01$. To further examine this complex three-way interaction, two separate follow-up ANOVAs of retention test data are reported below, one of initial lesson response data and one of correct response data. But first this three-way interaction is graphically compared to the delta rule predicted values (see Figure 4) to set the stage for the follow-up analyses.

Since the observed retention data shown in Figure 4 consists of multiple-choice questions, these data were corrected for guessing in order to be consistent with the predicted values. The correction-for-guessing formula from Nitko (1996) is corrected score = R - W / (n - 1) where R is raw score, W is number wrong, and n is the number of multiple-choice alternatives. The lines showing the predicted and observed retention for initial lesson responses (left panel of Figure 3) are highly similar in both magnitude and form, suggesting that the connectionist model presented by Clariana (1999a) may adequately account for 24-hour retention of initial learner responses. The lines showing the predicted and observed retention of correct responses (see right panel of Figure 4) are generally similar in shape but are not similar in magnitude. Specifically, the predicted correct response values overestimate the observed values.

FOLLOW-UP ANALYSIS OF RETENTION TEST ILR DATA

Retention test ILR data were analyzed by a mixed 3 × (2 × 3) ANOVA with one between-subjects factor, feedback condition (DF, MTF, or STF), and two within-subjects factors, kind of question (verbatim and inferential) and lesson item difficulty (most difficult, midrange, and easy). A significant effect was obtained for feedback, $F(2, 49) = 4.13$, $MSE = 0.081$, $p < .05$; as reported earlier, retention of ILRs was greater for DF compared to STF (see Figure 3); in addition MTF was more like STF than like DF.

A significant effect was also observed for kind of question $F(2, 49) = 8.24$, $MSE = 0.024$, $P < .01$ and for item difficulty blocks $F(2, 98) = 20.24$, $MSE = 0.025$, $p < .01$. Retention test memory of ILRs for verbatim questions ($M = 0.80$) was greater than memory of inferential questions ($M = 0.75$). Retention test memory of ILRs for difficult items (M = 0.72) and for middifficulty items (M = 0.76) was less than memory of easy items (M = 0.85). As indicated earlier, though these two findings are significant, they have little practical meaning.

FOLLOW-UP ANALYSIS OF RETENTION TEST CORRECT RESPONSE DATA

Retention test correct response data were analyzed by the same mixed 3 × (2 × 3) ANOVA. Three findings reached significance. A significant effect was obtained for item difficulty blocks, $F(2, 98) = 24.74$, $MSE = 0.025$, $p < .01$. Scheffé tests showed that retention test memory of correct responses for difficult items (M = 0.71) and for midrange items (M = 0.77) were both less than memory of easy items (M = 0.87), a finding of little practical interest.

More importantly, the interaction of feedback and item difficulty was significant, $F(4, 98) = 2.54$, $MSE = 0.025$, $p < .05$ (Greenhouse-Geisser $p = 0.05$; Huynh-Feldt $p = 0.04$). Although this disordinal interaction was directionally consistent with the delta rule predictions, with STF best for difficult items and DF best for easy items (see right panel of Figure 4), the follow-up Scheffé test indicated no significant differences for type of feedback within each question level. Thus the hypothesized differences between DF and STF at different item difficulty levels were too small to be considered reliable.

The interaction between kind of question and lesson item difficulty was significant, $F(2, 98) = 7.62$, $MSE = 0.023$, $p < .01$. Though inferential lesson questions appear to be slightly more effective than verbatim questions for difficult lesson items (see Figure 5), follow-up Scheffé test indicated no significant differences for type of feedback within each question level. Thus a possible instructional advantage of inferential questions over verbatim questions (per Merrill, 1987) is too small to be considered reliable.

PROBABILITY MEASURES

Previous feedback timing studies have examined the probabilities of failed and passed lesson items that are corrected (or not) on the retention test. Traditionally, these probabilities are calculated as test value divided by lesson value and are labeled as follows: R_2/R_2 — total items correct on the test divided by the total correct on the lesson, W_2/R_1 — total items incorrect on the test divided by those correct on the lesson, R_2/W_1 — total items correct on the test divided by those incorrect on the lesson, and W_2/W_1 — total items incorrect on the test divided by the those incorrect on the lesson. Probability data for this study are shown in Table 2.

Probability data for correct retention test responses were analyzed by a mixed 3 × 2 ANOVA with one between-subjects factor, feedback (DF, MTF, or STF), and one within-subjects factors, lesson correctness (R_2/R_1 and R_2/W_1). A significant effect was obtained for lesson correctness, $F(1, 49) = 70.12$, $MSE = 0.026$, $p < .01$, indicating that R_2/R_1 probability was greater than R_2/W_1 (see Figure 6). The interaction of feedback and lesson correctness was significant, $F(1, 49) = 3.14$, $MSE = 0.026$, $p < .05$. The means participating in this significant interaction are shown in Figure 6. The follow-up Scheffé test indicated no significant differences for type of feedback within either R_2/R_1 or R_2/W_1.

A closer look at these probability means shows that delayed feedback and immediate feedback obtain similar retention test scores through different processes. Specifically, DF maintains initial correct lesson responses while STF corrects lesson errors (relative). The STF probability of correcting lesson errors (R_2/W_1) was greater (approaching, but not significant) than that of MTF and DF, STF (.70) > MTF (.57) > DF (.53). The lower DF probability appears to be partly due to learners' repeating initially incorrect lesson responses on the retention test, which is called error perseveration. In this study, lesson errors were repeated on the retention test 28% of the time by the DF group but only 18% of the time by the STF group. This strong tendency of STF to correct errors in this study (R_2/W_1) should result in substantially greater posttest performance for the STF group. However such outcomes were negated by another effect, specifically, the probabilities of changing an initially correct lesson response to an incorrect response 24 hours later, W_2/R_1 (not significant): STF (.17) > MTF (.14) > DF (.11). Note that a similar result occurred in the study by Peeck and Tillema (1978), so this may be a common occurrence.

DISCUSSION

The first hypothesis was confirmed, that retention of initial lesson responses is greater for delayed feedback compared to immediate feedback across all item difficulties, but especially with difficult lesson items. With STF, retroactive interference at the instance of encoding is likely occurring, depressing memory of ILR errors relative to DF. In connectionist terminology, the association weights of ILR errors under STF decreased. With DF, lesson errors perseverated, or in connectionist terms, the association weights of ILR errors increased (see dashed line in Figure 3). Further, the observed retention of initial lesson responses for STF and DF were very similar to their corresponding delta rule predicted values (see left panel of Figure 4). This finding provides empirical support for the potential of a connectionist model to predict instructional feedback effects.

The practical value of Hypothesis One, for instructional design, is that in some learning situations, it is critical to remember initial lesson responses, especially if answers are not absolutely right or wrong but serve as learning transitions to broader understanding. For example, in discovery learning situations, learners are required to remember and use previous responses. In such cases, delayed feedback or even no feedback would allow learners to maintain initial lesson responses at a greater rate than with immediate feedback. On the other hand, immediate feedback involves a trade-off between increasing correct response associations at the expense of forgetting other responses, and these other responses are likely more meaningful to the learner even though incorrect. In the many situations where it is critical to strengthen the correct response and diminish the incorrect response, then immediate feedback would be better.

The second hypothesis, that feedback has its greatest effect with difficult lesson items, was confirmed. Lesson-to-retention test changes at each lesson item difficulty block expressed in effect sizes (ES), calculated as the difference between lesson and retention score divided by the standard deviation of the lesson score, are: for easy lesson items, ES = -.06; for midrange lesson items, ES = .35; and for difficult lesson items, ES = 1.17. Previous studies have provided the groundwork for this finding by showing that feedback has its greatest effect with difficult items (Sturges, 1978). For example, Bangert-Drowns et al. (1991) state, “If feedback’s primary importance is the correction of errors, then one would expect to see larger effects for instruction with higher error rates. This is exactly what happens” (p.230). Thus, future feedback investigations must consider and control lesson item difficulty, or else results may be confounded by lesson difficulty.

The third hypothesis, that feedback timing interacts with lesson item difficulty, was not supported. However the mean differences were in the right direction, with STF best with difficult items and DF best with easier items (see the right panel of Figure 4). This result could be anticipated in that lesson items were not difficult enough to produce the interaction; the average lesson item difficulty for the most difficult question block in this investigation was M = 0.50. Note that the delta-rule-predicted difference between immediate and delayed feedback would be most pronounced for lesson item difficulties less than about 0.40 (see Figure 2), and are actually predicted to be identical for lesson item difficulties near 0.50. Thus, to adequately test this hypothesis, an unusually difficult lesson would be required. Pragmatically, for computer-based lessons that use multiple-choice questions of reasonable difficulty, immediate and delayed feedback groups should obtain similar retention of correct responses, with perhaps a slight advantage with delayed feedback.

The results of this study related to Hypotheses One and Three tend to support the dual-trace explanation over the IPH. Hypothesis One was confirmed, that learners in the DF group remember their initial lesson responses quite well 24 hours after the lesson compared to the immediate feedback groups. This is the opposite of what is expected to occur with the IPH, but is consistent with the dual-trace explanation. Hypothesis Three, though not confirmed, obtained STF correct-response means that were greater than the DF correct-response means with difficult lesson items (see the right panel of Figure 4). This result is in the opposite direction of what should occur with the IPH. In summary, these findings are not potent enough to confirm or reject either explanation, but are more consistent with the dual-trace explanation than with the IPH.

What are the effects of feedback immediacy and of multiple exposures? MTF was much more like STF in retention test memory of ILRs, indicating that feedback immediacy acts to reduce memory of ILR errors, a retroactive interference effect. But MTF, rather than clearly mirroring STF (immediacy) or DF (multiple item exposure), generally fell midway between STF and MTF, indicating that both feedback timing and number of exposures combine or interact to impact retention test memory, especially for memory of correct responses. This combination or interaction of immediacy and multiple-exposures is of theoretical interest, and so should be addressed by additional experimentation.

With MTF, there is probably more involved than just immediacy and number of exposures. For example, Dick and Latta (1970) reported that low-ability learners receiving multiple-try constructed response feedback became frustrated when they made many lesson errors. In the same way, Clariana (1999b) reported that learners under MTF were more frustrated than those under STF, and warned against using MTF with difficult lesson materials. That MTF can be frustrating is supported somewhat by the slightly lower scores for MTF relative to STF with difficult items (see Figure 4). Thus instructional designers should consider the possible negative effects of requiring second and third tries with difficult lesson items.

The findings of this study involve only retention test recognition learning outcomes and should not be generalized to other types of learning outcomes, such as recall. Additional research should consider the application of a connectionist model for predicting the effects of feedback on higher-level learning outcomes.

As a footnote, Kulhavy and Stock (1989) have described an information processing explanation of feedback effects based on servocontrol theory (which describes the interaction between system output, sensors, feedback from the sensors, and mechanical devices that impact output). Their model views the learner's response confidence as a metacognitive component that controls or at least strongly influences how feedback information is processed by the learner. So far (since Sturges, 1978), response confidence studies have shown mixed results (Bangert-Drowns et al., 1991; Mory, 1992; 1994), so whether response confidence serves a metacognitive function in feedback processing is unknown. Requiring a learner to consciously consider the confidence of every lesson response would obviously alter the expected and normal pattern of a lesson, for example increasing the amount of time the learner takes with each item. Further, asking the learner for response confidence at each lesson response may be distracting, and could disrupt the learning process. Also, it has been suggested by Mory (1994) that learners' self-report of response confidence is inaccurate in some cases (see also Metcalfe, 1986). For these reasons, response confidence was not used in this present investigation. However, response confidence could serve as a logical alternative measure of initial lesson output activation level, a_{out}, in the delta rule calculations. Thus, future investigation of the possible metacognitive effects of response confidence may obtain added insight by applying a connectionist model.

ADDED MATERIAL

Roy Clariana is assistant professor in Instructional Systems at Penn State Great Valley in Malvern. His e-mail address is: RClariana‘psu.edu.

Daren Wagner and Lucia C. Roher Murphy are doctoral candidates in the instructional systems program at Penn State.

Table 1 * Lesson time data (in seconds), and lesson and retention test means and standard deviations (in parentheses) for each treatment group for each item kind and difficulty level

			Lesson		Retention Test			
Treatment	Lesson time	Item difficulty	V	I	ILR-V	ILR-I	CR-V	CR-I
STF (n=18)	1577	Difficult	0.46	0.44	0.66	0.66	0.67	0.83
	(280.6)		(0.20)	(0.10)	(0.18)	(0.26)	(0.22)	(0.18)
		Mid-range	0.85	0.57	0.76	0.67	0.79	0.67
			(0.07)	(0.13)	(0.18)	(0.19)	(0.18)	(0.21)
		Easy	0.92	0.82	0.87	0.78	0.91	0.84
			(0.08)	(0.11)	(0.12)	(0.18)	(0.09)	(0.12)
MTF (n=17)	1554	Difficult	0.51	0.55	0.71	0.67	0.71	0.72
	(332.5)		(0.20)	(0.16)	(0.24)	(0.17)	(0.20)	(0.21)
		Mid-range	0.77	0.65	0.79	0.72	0.81	0.74
			(0.07)	(0.04)	(0.18)	(0.16)	(0.17)	(0.17)
		Easy	0.91	0.83	0.86	0.81	0.91	0.81
			(0.09)	(0.09)	(0.12)	(0.18)	(0.12)	(0.18)
DF (n=17)	1899	Difficult	0.48	0.55	0.80	0.81	0.68	0.68
	(409.9)		(0.18)	(0.14)	(0.18)	(0.22)	(0.25)	(0.27)
		Mid-range	0.85	0.65	0.84	0.78	0.85	0.78
			(0.06)	(0.07)	(0.15)	(0.18)	(0.14)	(0.21)
		Easy	0.94	0.78	0.93	0.87	0.89	0.82
			(0.05)	(0.08)	(0.09)	(0.16)	(0.10)	(0.17)

Note: Lesson time is in seconds; standard deviations in parenthesis; V = Verbatim questions, I = Inferential questions; ILR = remembers initial lesson response; CR = recognize correct response; STF = Single-try feedback, MTF = Multiple-try feedback, DF = Delayed feedback. Each item difficulty range is the average of 6 items.

Table 2 * Observed retention test probabilities for passed and failed lesson items

	Passed lesson items	
Feedback	R2/R1	W2/R1
STF	0.84	0.17
MTF	0.86	0.14
DF	0.89	0.11
	Failed lesson items	
	R2/W1	W2/W1
STF	0.70	0.18(FN*) + 0.12(FN**)
MTF	0.57	0.22(FN*) + 0.21(FN**)
DF	0.53	0.28(FN*) + 0.19(FN**)

FOOTNOTES

* same error

** different error.

Figure 1 * A simple connectionist network with several possible output activation levels (a_{out}) for a given input pattern (a_{in}).

Figure 2 * Predicted retention test values generated by the delta rule across a range of lesson item difficulty values (from Clariana, 1999a). Predicted retention test memory of Initial Lesson Responses (ILR) are shown as dashed lines and Correct Responses (CR) are shown as solid lines for delayed feedback (DF) and immediate feedback (IF).

Figure 3 * Interaction of feedback and type of response which includes retention test memory of initial lesson response (ILR) and correct response (CR).

Figure 4 * Lesson Response and Correct Response predicted retention test values (dashed lines) for immediate feedback (IF) and delayed feedback (DF) compared to observed data corrected for guessing (solid lines) for delayed feedback (DF), multiple-try feedback (MTF), and single-try immediate feedback (STF).

Figure 5 * Comparison of inferential and verbatim question retention of correct responses across a range of lesson item difficulties. The dashed line represents the point where retention and lesson values are equal, the vertical arrows show lesson to retention test change by connecting mean lesson score to mean test score.

Figure 6 * Retention test correct response probabilities for each type of feedback for lesson errors (R_2/W_1) and for lesson correct responses (R_2/R_1).

REFERENCES

Ausubel, D.P. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston.

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(2), 213-238.

Brackbill, Y., Blobitt, W.E., Davlin, D., & Wagner, J.E. (1963). Amplitude of response and the delay-retention effect. Journal of Experimental Psychology, 66(1), 57-64.

Brown, J.I., Bennett, J.M., & Hanna, G. (1981). The Nelson-Denny Reading Test: Examiner's manual. Itasca, IL: Riverside Publishing.

Bruner, J. (1990). Acts of meaning. Cambridge, MA: Harvard University Press.

Clariana, R.B. (1999a). Differential memory effects for immediate and delayed feedback: A delta rule explanation offeedback timing effects. A Presentation at the Annual Convention of the Association for Educational Communications and Technology, February, 1999 in Houston, TX. (ERIC Document Reproduction Center ED 430 550)

Clariana, R.B. (1999b). CBT design: A feedback achievement interaction. Twenty-first Annual Proceedings of Selected Research Presentations of the Association for Educational Communications and Technology, 21, 87-92.

Clariana, R.B. (1993). A review of multiple-try feedback in traditional and computer-based instruction. Journal of Computer-Based Instruction, 20, 67-74.

Clariana, R.B., Ross, S.L., & Morrison, G.R. (1991). The effects of different feedback strategies using computer-assisted multiple-choice questions as instruction. Educational Technology Research and Development, 39(2), 5-17.

Crocker, L.M. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

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. Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99.

Evans, S.H., & Anastasio, E.J. (1968). Misuse of analysis of covariance when treatment effects and covariate are confounded. Psychological Bulletin, 69, 225-234.

Glover, J.A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. Journal of Educational Psychology, 81(3), 392-399.

Haberlandt, K. (1997). Cognitive psychology 2nd edition. Boston, MA: Allyn and Bacon.

Kennedy, J.J., & Bush, A.J. (1985). An introduction to the design and analysis of experiments in behavioral research. Lanham, MD: University Press of America.

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(4), 279-308.

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

Markowotz, N., & Renner, K.E. (1966). Feedback and the delay-retention effect. Journal of Experimental Psychology, 72(3), 452-455.

McLeod, P., Plunkett, K., & Rolls, E.T. (1998). Introduction to connectionist modeling of cognitive processes. Oxford, UK: Oxford University Press.

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

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. Journal of Experimental Psychology: Learning, Memory and Cognition, 12, 288-294.

Mory, E.H. (1992). The use of informational feedback in instruction: Implications for future research. Educational Technology Research and Development, 40(3), 5-20.

Mory, E.H. (1994). Adaptive feedback in computer-based instruction: Effects of response certitude on performance, feedback-study time, and efficiency. Journal of Educational Computing Research, 11(3), 263-290.

Nitko, A.J. (1996). Educational assessment of students, second edition. Englewood Cliffs, NJ: Prentice-Hall.

Peeck, J., & Tillema, H.H. (1978). Delay of feedback and retention of correct and incorrect responses. Journal of Experimental Education, 38, 171-178.

Plunkett K., & Marchman, V.A. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. Cognition, 48, 21-69.

Plunkett K., & Marchman, V.A. (1996). Learning from a connectionist model of the acquisition of the English past tense. Cognition, 61, 299-308.

Rankin, R.J., & Trepper, T. (1978). Retention and delay of feedback in a computer-assisted task. Journal of Experimental Education, 64, 67-70.

Sassenrath, J.M., & Yonge, G.D. (1969). Effects of delayed information feedback and feedback cues in learning on delayed retention. Journal of Educational Psychology, 50, 174-177.

Schroth, M.L., & Lund E. (1993). Role of delay of feedback on subsequent pattern recognition transfer tasks. Contemporary Educational Psychology, 18, 15-22.

Seidenberg, M., & McClelland, J. (1989). A distributed model of word recognition and naming. Psychological Review, 96, 523-568.

Shanks, D.R. (1995). The psychology of associative learning. Cambridge, UK: Cambridge University Press.

Shanks, D.R., Pearson, S.M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality. Quarterly Journal of Experimental Psychology, 41B, 139-159.

Sturges, P.T. (1969). Verbal retention as a function of the informativeness and delay of informative feedback. Journal of Educational Psychology, 60(1), 11-14.

Sturges, P.T. (1972). Information delay and retention: Effect of information in feedback and tests. Journal of Educational Psychology, 63(1), 32-43.

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

Sturges, P.T., Sarafino, E.P., & Donaldson, P.L. (1968). Delay-retention effect and information feedback. Journal of Experimental Psychology, 78(2), 357-358.

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

Swindell, L.K., & Walls, W.F. (1993). Response confidence and the delay retention effect. Contemporary Educational Psychology, 18, 363-375.

SYSTAT 8.0. [Computer software]. (1998). Evanston, IL: SPSS Inc.

Widrow, B., & Hoff, M.E. (1960). Adaptive switching circuits. 1960 IRE WESCON Convention Record (Pt. 4), 96-104.