DOCUMENT RESUME

ED 360 332 TM 020 129

AUTHOR Bender, Timothy A.
TITLE Predicting Postfeedback Performance from Students' Confidence in Their Responses.
PUB DATE Apr 93
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Comparative Analysis; Confidence Testing; Correlation; Educational Assessment; *Feedback; Higher Education; Knowledge Level; Models; *Performance; *Predictive Measurement; Regression (Statistics); *Responses; Scores; Self Concept; Student Interests; Student Reaction; Test Use; *Undergraduate Students
IDENTIFIERS Variance (Statistical)

ABSTRACT

The model of feedback processing proposed by R. W. Kulhavy and W. A. Stock (1989) was studied in a traditional classroom setting in which methods of assessing students' response confidence as predictors of postfeedback performance were also examined. The relationship between confidence ratings at the time of the test and confidence assessed prior to delayed feedback was explored. Subjects were 6 male and 21 female undergraduates assigned to confidence or interest conditions who rated their confidence in responses or interest in each questions. Data for 25 students were used. Students were asked to predict their scores, received feedback, and completed the examination again. An analysis of variance compared the performance of the confidence and interest groups, and regression and correlation analyses explored the predictability of postfeedback performance. There were no significant differences between postfeedback performance of the interest and confidence groups. Increase in elaborative processing due to students' rating their confidence does not appear to affect postfeedback performance any more than does rating the interest level. Results indicate that the Kulhavy and Stock model can be applied to the classroom. Use of students' estimates of test scores is not recommended as a measure of response confidence, as it accounted for very little variance in postfeedback performance. Implications for prediction of students' feeling-of-knowing are explored. (SLD)

Predicting Postfeedback Performance from Students' Confidence in Their Responses.

Timothy A. Bender
Southwest Missouri State University

Theoretical Framework and Objectives

Much of the current research involving exam feedback is not conducted in traditional classroom situations. For example, Kulhavy and Stock's (1989) model of feedback processing is based on programmed, computer-assisted instruction and testing. The typical paradigm for this research employs computer-assisted testing and feedback in which students read a passage, answer a question, record their confidence in their response, and receive immediate feedback. This differs from a traditional classroom situation in which the studying of the material, the testing, and the feedback are all separated by hours or days. One purpose of this research was to study the Kulhavy and Stock (1989) model in a traditional classroom setting.

The second purpose was to assess three means of measuring students' response confidence on a classroom exam as predictors of postfeedback performance on the same exam. A key component of the Kulhavy and Stock model is students' confidence in their responses. Students' response confidence affects their subsequent processing of the feedback for each response. Kulhavy and Stock (1989) assess response confidence by asking students to rate their confidence in each response on a 5-point scale. These responses are then translated into discrepancy scores. High discrepancy equates to high confidence on a wrong response; low discrepancy reflects high confidence on a correct response. Nelson (1984) favors a different method of assessing response confidence, ie. the Goodman-Kruskal gamma correlation, which reflects the accuracy of students' feeling-of-knowing (FOK). Finally, confidence may be reflected in students' predictions of their exam scores.

In order to assess the independent relationship between these measures of confidence and postfeedback performance, the effect of differences in knowledge of the tested material should first be removed. This is particularly true when mean discrepancy is used to reflect the overall response confidence for a test rather than for each test question. Discrepancy, by definition, must be negatively correlated with initial performance. The formula for determining discrepancy assigns negative values to the confidence levels of all correct responses and positive values to the confidence levels of all incorrect responses (Kulhavy & Stock, 1989). Thus, higher scores must be accompanied by lower discrepancy. On the other hand, the formula for determining gamma does not necessitate a relationship between gamma and initial performance (Nelson, 1984.)

The third purpose was to explore the relationship between students' confidence ratings at the time of the test with confidence assessed prior to delayed feedback. In the Kulhavy and Stock model, students rate their confidence and receive immediate feedback. Thus, confidence at the time of the initial response and confidence at the time of feedback are the same. However, in a traditional classroom setting, feedback is often delayed. Students' confidence at the time of receiving feedback may be different from that at the time of the initial response.

The final purpose was to discover if recording of response confidence affected students' initial and postfeedback test performance. Asking students to rate their confidence on each response may require greater elaboration of the test questions and material than simply responding to the questions. Also, rating the response confidence on each question provides the students with additional exposure to the information. In order to examine whether increased elaboration through forming of confidence ratings affected postfeedback performance, half of the students were asked to provide confidence ratings and the other half to rate the test questions for their interest value. I assumed rating the interest value of questions would produce less elaboration on the content of the question, but maintain the amount of exposure to the questions.

Method

Subjects

Subjects were 6 male and 21 female undergraduates enrolled in my evening section of Educational Psychology. Because 2 students did not complete the exams at the regularly scheduled time, only the data from the remaining 25 students were used.

Procedure

The class met once a week for three hours per week. At the beginning of the fifth, tenth and sixteenth week the class was administered a 50-question, multiple-choice exam covering that one-third of the material. At the start of the class meeting following the first and second exams, the exams were returned with the number of correct responses marked at the top of the answer sheets. After writing the grade distribution on the blackboard, I read each question to the students and explained the reasoning that supported the correct answer. After reviewing the exam, I answered any questions about any test question. I then collected all exams and answer sheets.

Prior to the administration of Exam Two, a matching procedure based on students' performance on Exam One was used to assign students to confidence or interest conditions. After students completed Exam Two, they rated either their confidence in each response or interest in each question, using a 5-point scale. Directions for these ratings were provided on the tests, ensuring that students did not know they were in different conditions. As they returned the exams and answer sheets, students in both conditions were asked to predict their scores.

Before receiving feedback about their performance on Exam Two, students were handed the exams, asked to re-read the questions, and requested to once more rate the questions. Students in the confidence condition rated their current confidence; and those in the interest condition rated their current interest. As the students returned the rating sheets, they were again asked to predict their scores. Once the ratings were completed, feedback was provided as described above. Following the feedback session, the exams and answer sheets were collected and the exam was re-administered. Students were told they would receive the higher of the two scores. The students were not told that they would be repeating the exam.

Results

Data were analyzed using simple one-way ANOVAs to compare the performance of the confidence and interest groups and by regression and correlational analyses to explore the predictability of postfeedback performance. Dependent variables included the raw scores on the second administration of Exam Two, the conditional probability of correcting an error, and the conditional probabilities of committing perseverative, different, and new errors (Phye and Bender, 1989).

Stability of Confidence Measures

Gamma, discrepancy, and the difference between the achieved and predicted scores were used to assess students' response confidence. These scores were determined both at the time of testing and at the feedback session one week later. As seen in Table One all three measures were

Table 1

Correlations of Confidence at Test and Feedback

Confidence Measure	Correlation	p value
Discrepancy (n = 13)	0.977	.00009
Gamma (n = 13)	0.846	.0003
Difference Score (n = 24)	0.693	.0002

significantly ($p < .05$), positively related across time, with correlations of .977 for discrepancy, .846 for *gamma*, and .693 for difference scores.

Confidence Measures and Initial Test Scores

All three confidence measures were significantly ($p < .009$) related to initial scores on Exam Two. As can be seen in Table Two, students who scored well on the first administration of the exam performed better than they expected, were more accurate in their feelings of knowing the material (*gamma*), and experienced less discrepancy.

Table 2

Correlations of Confidence Measures and Initial Test Performance

Confidence Measures at Time of Test	Correlation	p value
Discrepancy (n = 13)	-0.979	.00009
Gamma (n = 13)	0.862	.0002
Difference Score (n = 13)	0.516	.009
Confidence Measures at Feedback		
Discrepancy (n = 13)	-0.974	.00009
Gamma (n = 13)	0.787	.0002
Difference Score (n = 13)	0.661	.009

Predicting Postfeedback Performance

Regressions and partial correlations were used to determine the extent to which each confidence measure individually predicted postfeedback performance, independent of initial performance on the exam. Table Three presents the partial correlations from all significant regressions.

Table 3

Partial Correlations for Gamma and Discrepancy with Posttest Score and Probability of New Errors

Regressors for Posttest Score	Partial r	Partial r^2
Initial Test Score	-0.082	0.007
Gamma at Test	0.505	0.255
Regressors for Probability of New Errors		
Initial Test Score	0.369	0.136
Discrepancy at Test	0.498	0.248
Initial Test Score	0.128	0.016
Gamma at Test	-0.544	0.296

The regression of posttest scores onto initial performance and the difference between the achieved and expected scores for Exam Two, when determined at the time of the test was significant $F(2,22) = 10.7$, $p < .001$, adjusted $R^2 = .447$. Initial performance accounted for approximately one-third of the variance in posttest scores. However, independent of initial performance, differences between achieved and predicted scores on the initial test accounted for less than 6 percent of the posttest variance.

Similar results were found when posttest scores were regressed onto initial performance and the difference between the expected and achieved scores, assessed just prior to the feedback, $F(2,21) = 7.93$, $p < .003$, adjusted $R^2 = .376$. Initial performance accounted for 34% of the variance. The difference between expected scores and achieved scores accounted for less than 2% of the variance.

The regression of posttest scores onto initial performance and gamma based on confidence at the time of testing was also significant $F(2,10) = 5.17$, $p < .03$, adjusted $R^2 = .410$. Independent of gamma, initial performance accounted for under 1 percent of the variance in posttest scores. On the other hand students' feeling-of-knowing (gamma), independent of initial exam performance, was a strong, positive predictor of posttest performance, accounting for 25% of the variance.

The regression of the probability of committing a new error onto initial performance and the difference between the achieved and predicted scores, at the time of testing, was significant $F(2,22) = 4.44$, $p < .03$, adjusted $R^2 = .223$. Initial performance was negatively related to committing new errors, accounting for 26% of the variance. The difference between the expected and achieved scores was positively related to committing new errors, but accounted for less than 2% of the variance.

Results of the regression of the probability of new errors onto initial performance and the difference between achieved and expected scores, when the prediction was made just prior to feedback were similar to the aforementioned regression, $F(2,21) = 4.26$, $p < .03$, adjusted $R^2 = .221$.

The regression of the probability of committing a new error onto initial performance and discrepancy at the time of testing was significant $F(2,10) = 5.01$, $p < .04$, adjusted $R^2 = .40$. Initial performance, independent of discrepancy, accounted for approximately 14% of the variance of the probability of committing new errors. Discrepancy was positively related to the probability of committing new errors, accounting for about 25% of the variance.

The regression of the probability of committing new errors onto initial performance and *gamma* at the time of testing was also significant $F(2,10) = 5.70$, $p < .03$, adjusted $R^2 = .44$. Initial performance accounted for less than 2% of the variance. However, *gamma* was related negatively to committing new errors, accounting for almost 30% of the variance.

Confidence and Elaboration

There were no significant differences between the postfeedback performances of the interest and confidence groups. Thus, any increase in elaborative processing of the test questions due to students rating their confidence in their responses does not appear to affect postfeedback performance any more than the less elaborative task of rating their interest in the test questions. Alternatively, there may be no difference in elaborative value between rating ~~test~~ the interest level of questions versus response confidence.

Discussion

The Kulhavy and Stock (1989) model of feedback processing can be applied to the classroom setting. However, I propose some recommendations concerning the measurement of students' response confidence. First, if researchers are interested in the unique relationship between students' confidence and their use of feedback, the effects of initial test performance should first be removed. Second, the method used to assess students' response confidence should be carefully considered, as each method has its advantages and disadvantages. Third, response confidence should be assessed at the time the responses are made, rather than at the time feedback is provided.

Predicted Scores

The use of students' estimates of their test scores is not recommended as a measure of their response confidence. It has the advantages of being the easiest to obtain and may be used to determine the difference between the expected and attained score. However, it accounted for very little variance in postfeedback performance. Predicted test scores may not be sensitive to the impact of the confidence associated with each individual response.

Discrepancy

One of the advantages of discrepancy is that it can be applied to students' processing of individual items. Kulhavy and Stock (1989) found that discrepancy is related to the time students spend in reviewing feedback on individual test questions. Also, it is a relatively simple value to obtain, requiring only a simple rating scale and minor calculations.

I recommend that the data for determining discrepancy be obtained at the time the students respond to the test rather than just before the feedback. While discrepancy scores were very consistent across time, those obtained at the time of testing were the best predictors of postfeedback performance. Students who experienced a higher degree of discrepancy were more likely to commit a new error. Apparently, during feedback these students focused on those questions they missed and did not take the opportunity to confirm the questions they had answered correctly.

Discrepancy has the disadvantage of being strongly related to scores on the initial test. When the mean of the discrepancies on individual items is used as a test-wide measure of confidence, a strong negative correlation appears between discrepancy and initial test performance. Therefore, it is important to remove the effects of the initial test score if discrepancy is used as a test-wide measure of confidence.

Gamma

The Goodman-Kruskal gamma correlation has been previously recommended over other measures of the feeling-of-knowing (Nelson, 1984). That recommendation is echoed here. Gamma shares many characteristics with discrepancy. First, it is not difficult to determine, requiring only a rating scale and simple calculations (Nelson, 1984). Second, gamma is stable across time. Better than 70% of the variance was shared between gamma at the time of testing and that obtained just prior to feedback. Third, gamma also should be determined at the time of the initial test. Fourth, although the formula for gamma does not force a relationship with initial test performance, such a relationship should be investigated and controlled.

Unlike discrepancy, Gamma accounted for variance in posttest scores (25%) and the probability of committing new errors (30%.) Independent of initial test scores, students who were more accurate in their feelings-of-knowing the material scored well on the posttest and were not likely to commit new errors. This low probability of committing new errors supports the validity of gamma as a measure of the feeling-of-knowing.

Limitations of the Study

My recommendations are only tentative, as the study suffers from limitations common to much educational research conducted in vivo, that of a small sample size. Although a matching procedure was used to form the confidence and interest conditions, individual subject characteristics still may have had a large impact on the results. Also, due to the small sample size, a no-treatment control group was lacking. A control group which did not report confidence or interest would have provided information about whether recording response confidence then getting feedback affects later performance differently than simply receiving feedback.

A small sample size only decreases the likelihood of find a significant difference. It does not increase the likelihood of committing the Type 1 error of falsely rejecting a true null. Therefore, the finding of significant effects argues that confidence is important in students' use of feedback. Better-constructed research is needed to further explore the effect of response confidence in classroom exam settings.

Conclusion

In conclusion, the Kulhavy and Stock (1989) model of feedback processing can be applied to the classroom. Students' confidence in their responses is predictive of postfeedback performance and assessment of this confidence should occur at the time of testing.

The results of this study also support Nelson's (1984) recommendations of the Goodman-Kruskal gamma correlation as the preferred measure of the accuracy of students' feeling-of-knowing. Gamma was a significant predictor of both posttest scores and the probability of committing new errors. Discrepancy predicted only new errors. Like discrepancy, gamma was strongly correlated with initial performance, but this relationship was not necessitated by the formula for determining gamma.

Both gamma and discrepancy may be important in understanding the effects of feedback on postfeedback performance, as they reflect different aspects of students' confidence. Gamma reflects the accuracy of the feeling-of-knowing, while discrepancy reflects students' reactions to differences between the feedback and response confidence on individual test questions.

References

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

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95, 109-133.

Phye, G. D. & Bender, T. A. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. Contemporary Educational Psychology, 14, 97-110.