DOCUMENT RESUME

ED 073 156 TM 002 403

AUTHOR Kaufman, David; Conry, Robert
TITLE Evaluation of a Computer-Based Feedback System.
PUB DATE 30 May 72
NOTE 15p.; Paper presented at the Learned Societies Conference, Montreal, May 30, 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Analysis of Variance; College Students; *Computer Programs; *Feedback; *Grade Prediction; Higher Education; *Predictive Validity; *Program Evaluation; Student Attitudes; Test Results

ABSTRACT

This paper evaluates the effects of two distinct modes of feedback to students. These two modes comprise: (1) predicted final course grade, and (2) current standing in class. The experiment was conducted with a final year class of 56 students attending a required half-year course in Educational Measurement. The class was divided into four treatment groups (2 x 2 factorial design) and different feedback of results was mailed to each group for the four objective class tests written before final exam. Two FORTRAN computer programs were used to determine class standing, current grades and predicted final course grades. The predicted final exam and course grades were obtained by applying a linear least-squares line of best fit to the test data and extrapolating to future tests and to the final exam. The class standing was obtained by using a FORTRAN sorting program. Analysis of variance was performed on the final exam grades, and no significant differences were found in final exam performance or in satisfaction with feedback received (Chi-squared test), although results slightly favored the feedback groups. The computer program was found to adequately predict the final exam grades, i.e., the linear regression was significant. (Author/DB)

TM

EVALUATION OF A COMPUTER-BASED FEEDBACK SYSTEM

A paper presented by

David Kaufman
Faculty of Education, U.B.C.

and

Robert Conry
Faculty of Education, U.B.C.

at the

Learned Societies Conference
McGill University, Montreal
May 30, 1972

EVALUATION OF A COMPUTER-BASED FEEDBACK SYSTEM

ABSTRACT

This paper evaluates the effects of two distinct modes of feedback to students. These two modes comprise: (1) predicted final course grade, (2) current standing in class.

The experiment was conducted with a final year class of fifty-six students attending a required half year course in Educational Measurement. The class was divided into four treatment groups (2 x 2 factorial design) and different feedback of results was mailed to each group for the four objective class tests written before final exam.

Two FORTRAN computer programs were used to determine class standing, current grades and predicted final course grades. The predicted final exam and course grades were obtained by applying a linear least-squares line of best fit to the test data and extrapolating to future tests and to the final exam. The final course grade is simply a summed weighted composite of the class tests and final exam. The class standing was obtained by using a FORTRAN sorting program.

Analysis of variance was performed on the final exam grades and no significant differences were found in final exam performance or in satisfaction with feedback received (Chi-squared test), although results slightly favoured the feedback groups. The computer program was found to adequately predict the final exam grades, i.e., the linear regression was significant.

EVALUATION OF A COMPUTER-BASED FEEDBACK SYSTEM

Introduction

Ley (1) has summarized the advantages of short, frequent quizzes as leading to increased motivation, elimination of cremming, provision of continuous knowledge of student progress and removal of the fear of final examinations. He has developed a FORTRAN computer program which can be used for calculating student grades as well as for predicting the final exam and course grades. A program has been written by the authors to rank order the students after each test.

The purpose of this experiment was to evaluate the accuracy of the program as a predictor of final exam grades and to determine whether or not different modes of feedback from class tests resulted in significantly different final exam performance. The students' satisfaction with the different modes of feedback was also tested.

Design

Fifty-six final year Education students attending a 1½ unit course in Measurement at the University of British Columbia were used as subjects in the experiment. The students were randomly assigned to one of four treatment conditions as shown in Figure 1. The unequal cell sizes were caused by students withdrawing from the course. Feedback Condition A will be referred to as Norm-Referenced, and consisted of information regarding current class standing. Feedback condition B, referred to as Criterion-Referenced, consists of information

1. Ley, James B. "Grades, Quizzes, Motivation, and Computers", IEEE Transactions on Education, Vol. E-13, No.3, September, 1970.

about predicted final course grade. The feedback given to each group is summarized in Table 1.

		B (Criterion) yes	B (Criterion) no
A (Norm)	yes	n = 15	n = 14
	no	n = 14	n = 13

Table 1 - Summary of Feedback Conditions

Feedback Mode		Information given to students		
A	B	Average % of correct responses	Class standing (which quarter)	Predicted final course grade
yes	yes	x	x	x
yes	no	x	x	
no	yes	x		x
no	no	x		

In order to minimize "leakage" between groups, the results were mailed to the students' homes after each quiz. Questions by some students about differences in the feedback were answered discretely by telling the students that the instructor wished to determine the feedback most satisfactory to them.

Four tests and a final examination were developed and administered in class by Prof. R. Conry, the instructor for the course. The tests were approximately one half-hour long and one hour was allotted to the final examination. The inter-test correlations are given in Table 2.

Table 2 - Test and Exam Correlations

	T1	T2	T3	T4	F. Exam
T1	1.000	.423	.270	.437	.412
T2		1.000	.169	.188	.296
T3			1.000	.347	.318
T4				1.000	.231
F. Exam					1.000

Computer Program

The computer programs used to obtain the feedback are given below. It should be noted that only certain parts of the overall output were revealed to each group.

```
      C     CALCULATION AND PREDICTION OF COURSE GRADES
      C
001         WRIT (6,33)
002      33 FORM T(1H1,20X,'CLASS MARKS'///)
      C
      C
003         DIMENSION ANAME(20),G(19)
      C
      C     THE INPUT DATA IS READ INTO THE COMPUTER
      C
004      100 READ(5,1,END=200)(ANAME(I),I=1,20),FEXAM,(G(J),J=1,19)
005        1 FORMAT(20A1,F3.0,19(F3.0))
      C
      C     THE CALCULATION OF THE AVERAGE GRADE
      C
006         GSUM=0.0
007         DO 10 I=1,19
008         IF(G(I))10,20,10
009      20 K=I-1
010         GO TO 15
011      10 GSUM=GSUM+G(I)
012      15 AVE=GSUM/K
      C
      C     THE TEST TO SEE WHETHER THE FINAL EXAM GRADE IS RECORDED
      C
013         IF(FEXAM)30,3 ,40
      C
      C     LINEAR LEAST SQUARES PREDICTION OF THE FINAL EXAM GRADE.
      C
014      30 SUM1=K
015         SUM2=K*(K+1)/2
016         SUM3=0.0
017         SUM4=GSUM
018         SUM5=0.0
019         DO 50 I=1,K
020         SUM3=SUM3+I*I
021      50 SUM5=SUM5+G(I)*I
022         DEN=SUM1*SUM3-SUM2*SUM2
023         A=(SUM1*SUM5-SUM2*SUM4)/DEN
024         B=(SUM1*SUM4-SUM2*SUM5)/DEN
025         FEXAM=A+B*20
      C
      C     THE WEIGHTED CALCULATION OF THE FINAL COURSE GRADE.
      C
026      40 FGRADE=0.4*AVE+0.6*FEXAM
      C
      C     THE COURSE RESULTS ARE PRINTED.
      C
027         WRITE(6,2)(ANAME(I),I=1,20),K,AVE,FEXAM,FGRADE
028      2 FORMAT(1H ,20A1,' AVERAGE OF',I4,' GRADES =',F5.1, /,' ESTIMAT
          1D FINAL EXAM GRADE =',F5.1,' ESTIMATED FINAL COURSE GRADE =',F5.1
          2,//)
029         GO TO 100
030      200 STOP
031         END
```

CLASS MARKS — SAMPLE OUTPUT

1 [redacted] 1 7551 AVERAGE OF 4 GRADES = 65.0
ESTIMATED FINAL EXAM GRADE = 62.0 ESTIMATED FINAL COURSE GRADE = 63.2

1 [redacted] 2 7532 AVERAGE OF 4 GRADES = 62.8
ESTIMATED FINAL EXAM GRADE = 55.0 ESTIMATED FINAL COURSE GRADE = 58.1

1 [redacted] 2 8556 AVERAGE OF 4 GRADES = 73.3
ESTIMATED FINAL EXAM GRADE = 72.0 ESTIMATED FINAL COURSE GRADE = 72.5

1 [redacted] 2 14 AVERAGE OF 4 GRADES = 56.8
ESTIMATED FINAL EXAM GRADE = 60.0 ESTIMATED FINAL COURSE GRADE = 58.7

1 [redacted] 2 8051 AVERAGE OF 4 GRADES = 56.3
ESTIMATED FINAL EXAM GRADE = 78.0 ESTIMATED FINAL COURSE GRADE = 69.3

1 [redacted] 2 [illegible]2 AVERAGE OF 4 GRADES = 7[illegible].5
ESTIMATED FINAL EXAM GRADE = 72.0 ESTIMATED FINAL COURSE GRADE = 74.6

STOP 0
EXECUTION TERMINATED

$SIG

FORTRAN IV G COMPILER MAIN 05-09-72 17:14:07 PAGE 000

```
C  PROGRAM TO RANK ORDER STUDENTS
C
C DATA CARDS ARE 1. FORMAT CARD OF FORM (20A1,F3.0)
C                2. NUMBER OF STUDENTS(COL. 1-3)
C                3. DATA CONSISTING OF ID CODE(COL. 1-20) AND MARKS
C
001          DIMENSION FMT(20),NAME(200,20),GRADE(200),TNAM(20)
002          READ(5,1)FMT
003        1 FORMAT(20A4)
004          READ(5,2)N
005        2 FORMAT(I3)
006          DO 3 M=1,N
007        3 READ(5,FMT)(NAME(M,J),J=1,20),GRADE(M)
008          N1=N-1
009          DO 11 I=1,N1
010          I1=I+1
011          DO 11 J=I1,N
012          IF(GRADE(I)-GRADE(J))14,11,11
013       14 TEMP=GRADE(I)
014          GRADE(I)=GRADE(J)
015          GRADE(J)=TEMP
016          DO 2 L=1,20
017          TNAM(L)=NAME(I,L)
018          NAME(I,L)=NAME(J,L)
019       20 NAME(J,L)=TNAM(L)
020       11 CONTINUE
021          WRITE(6,4)
022        4 FORMAT(1H1,10X,'RANK ORDER'///)
023          DO 30 M=1,N
024       30 WRITE(6,7)(NAME(M,I),I=1,20),GRADE(M),M
025        7 FORMAT(1H ,20A1,F5.0,I5)
026          STOP
027          END
```

TOTAL MEMORY REQUIREMENTS 0 46CA BYTES

COMPILE TIME = .4 SECONDS

RANK ORDER

SAMPLE OUTPUT

1	[redacted]	2	8051 78.	1
1		2	3556 72.	2
1		2	7032 72.	3
1		1	1551 62.	4
1		2	14 60.	5
1		2	1532 55.	6

STOP 0
EXECUTION TERMINATED

$SIG

Analysis and Results

All analyses were carried out at the UBC Computing Centre using an IBM 360/67 Computer.

The programs used in the analyses are summarized in Table 3.

Table 3 - Computer Programs

Analysis	Program
Analysis of Variance	BMDX64*
Regression Analysis	BMDO2R*
Chi - Square	MVTAB**

* The BMD programs are from the UCLA Biomedical Package of Statistical routines.

** MVTAB is a UBC computing centre program.

Analysis of variance on the final exam grades was performed (Table 4b) to test for differences in performance between the cell means (Table 4a) of the four groups.

Table 4a - Cell Means for Final Exam

A \ B	yes	no
yes	67.5	72.3
no	65.1	65.9

Table 4b - ANOVA

Source of Variation	df	ms	F	P <
A	1	271.86	1.81	.17
B	1	109.64	.73	.40
A X B	1	53.12	.35	.56
ERROR	52	150.24		

Stepwise regression analysis was performed using the final exam grade as the predicted variable and the test grades as predictor variables (Table 5a). The significance of the linear regression used in the predicting program was tested. (Table 5b)

Table 5a - Stepwise Regression

Step No.	Variable Entered	Multiple R	R^2	Increase in R^2	F value to enter	No. of ind. variables included
1	Test 1(T1)	.423	.179	.179	11.75	1
2	Test 2(T2)	.468	.219	.041	2.75	2
3	Test 3(T3)	.565	.319	.099	7.59	3
4	Test 4(T4)	.607	.368	.049	3.99	4

F. Exam = 27.7 + .195 (T1) + .027 (T2) + .233 (T3) + .156 (T4)

Table 5b - Anova for Regression

Source of Variation	df	ms	F	P<
Regression	4	549.27	7.43	.01
Residual	51	73.94		

A Chi - Square analysis was performed on the bivariate frequency tables of feedback condition vs. satisfaction with feedback received (yes or no). The analysis was carried out on an overall basis (Table 6a) and separately for the two feedback modes (6b, 6c).

Table 6a - Bivariate Table of Feedback vs Satisfaction

Feedback mode	Satisfaction: No	Satisfaction: Yes	
A and B	4	10	14
A only	6	8	14
B only	8	5	13
Neither	7	6	13
	25	29	54

$X^2_{(3)} = 3.35$ $P < .34$

Table 6b - Feedback vs. Satisfaction

Class standing Feedback (A) given	Satisfaction: No	Satisfaction: Yes	
No	15	11	26
Yes	10	18	28
	25	29	54

$X^2_{(1)} = 1.81$ $P < .17$

Table 6c - Feedback vs. Satisfaction

Predicted grade Feedback (B) given	Satisfaction: No	Satisfaction: Yes	
No	13	14	27
Yes	12	15	27
	25	29	54

$X^2_{(1)} = .00$ $P < .95$

Conclusions

The analysis indicates that there were no significant differences in final exam performance and in satisfaction between the groups. The group receiving feedback class standing does appear to be superior on both these variables but the small sample used in this experiment would require large differences in order to reach a significance level of $\alpha = .05$

There is a significant linear relationship between test grades and final exam grade, therefore justifying the use of the program to predict final grades. The linear regression accounted for 37% of the variance in final exam results.

Discussion

The students were asked for written comments about the feedback during the final examination. An apparent misunderstanding of the nature of feedback type B (predicted final mark) was evident. This was difficult to avoid because of the necessity of providing four different modes of feedback to a single class and attempting to avoid "leakage" between groups. This mode of feedback was earlier found by the author to be extremely popular with second-year Engineering students when the results for the whole class were regularly posted and the method of obtaining the feedback was explained.

The comments indicated a favorable response to feedback mode A (class standing), but more specific information such as rank or standard score based on class mean and standard deviation would have been more satisfactory. The choice of quarters (not quartiles) as an indicator of class standing was made in order to minimize "leakage" between groups, i.e., a student could not determine his rank by using information obtained from another student.

It should be noted that the subjects were in their final semester of university and other factors such as concern with employment may have reduced the impact of the feedback. Perhaps the regularity and not simply the mode of the feedback was the important factor for all the groups. Even under the restricted conditions of the experiment, the results showed a slight advantage for the feedback groups. It is hypothesized that a similar experiment performed with several separate classes of lower year students would yield more definite results.

It would be useful to determine whether a higher degree relationship exists between test grades and final exam grade, e.g.,quadratic, which would

account for a higher proportion of the total variance. If such a relationship exists, the FORTRAN predicting program should be modified accordingly.

Although the different feedback modes caused no change in performance, the predicting program can effectively be used to monitor the progress of students in a course. The "predicted final grade" would aid the instructor in identifying potential failures in time to provide the required individual assistance.