DOCUMENT RESUME

ED 205 554	TM 810 433
AUTHOR	Bliss, Leonard B.
TITLE	An Empirical Test of a Strategy for Training Examinees in the Use of Partial Information in Taking Multiple Choice Tests.
PUB DATE	13 Apr 81
NOTE	23p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Algorithms; Guessing (Tests); Intermediate Grades; *Multiple Choice Tests; Pretests Posttests; Scoring Formulas; *Test Coaching; Training Methods; True Scores
IDENTIFIERS	*Correction for Guessing; Iowa Tests of Basic Skills; *Number Right Scoring; Virgin Islands (Saint Thomas)

ABSTRACT

The aim of this study was to show that the superiority of corrected-for-guessing scores over number right scores as true score estimates depends on the ability of examinees to recognize situations where they can eliminate one or more alternatives as incorrect and to omit items where they would only be guessing randomly. Previous investigations show examinees lack this ability. An instructional strategy for teaching examinees to use partial information in taking multiple choice tests under formula scoring conditions was tested on 280 fourth, fifth and sixth grade, private and public school students from the United States Virgin Islands. This method is outlined and followed by sections dealing with the administration of pre-tests and post-tests and the provision of training to selected groups (in the form of an algorithm for attacking test items, and practice sessions with immediate feedback). It is indicated in the final discussion of research results that the technique was successful in training examinees to omit only items for which they had no more than a chance probability of correctly responding. References, figures of the algorithm and pre-test practice items, and tables of descriptive statistics are appended. (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED205554

An Empirical Test of a Strategy for Training Examinees in the Use of Partial Information in Taking Multiple Choice Tests

Leonard B. Bliss
Caribbean Research Institute
College of the Virgin Islands

Running head: Partial information

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Presented at the Annual Meeting of the American Educational Research Association, Los Angeles, April 13, 1981.

TM 810403

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. Bliss
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

The superiority of corrected-for-guessing scores over number right scores as true score estimates depends on the ability of examinees to recognize situations where they can eliminate one or more alternatives as incorrect and to omit items where they would only be guessing randomly. Previous investigations show examinees lack this ability. This study tested an instructional strategy for teaching examinees to use partial information in taking multiple choice tests under formula scoring conditions. The results indicated that the technique was successful in training examinees to omit only items for which they had no more than a chance probability of correctly responding.

The purpose of this study was to test the effectiveness of a training strategy designed to teach students to use partial information which they may have at their disposal when answering items on multiple choice tests.

Lord (1975) showed that using the correction for guessing in administering and scoring multiple choice tests would provide scores which would have higher levels of reliability and validity than scores obtained using number right directions. Number right procedures result in a score which is the actual number of items the examinee marks correctly, while the formula score is this number of correct items minus a fraction of the number marked incorrectly (usually W/C-1, where W is the number marked wrong and C is the number of choices per item). In the former situation examinees are directed to answer all items while in the latter they are directed to leave an item blank if they can make no better than a random guess at the answer. The superiority of the formula scoring method is dependent on the assumption that the examinees follow these directions and omit only items for which they would have to guess from among all choices. Examinees who could eliminate one or more choices as incorrect (i.e. had some partial information concerning the item) should guess from among the remaining choices.

However, numerous studies including those by Sheriffs and Boomer (1954), Crehan, Candor, & Beckett (1976), Slakter, Juliano, & Sarnacki (1976), Cross and Frary (1977), Poggio, Amus, & Levy (1978), and Bliss (1980) have shown that examinees omit items that they had a better than chance probability of answering correctly and were, therefore, penalized by the use of the correction for guessing formula. This phenomenon has been observed across grade levels from elementary school through university levels. Further, Bliss noted that, in the intermediate grades (4 through 6) it was the higher achieving students who tended to report having followed formula scoring directions and were, therefore, penalized most often.

It is possible that formula scoring directions (answer if you know <u>anything</u>, omit if you know <u>nothing</u>) make unreasonable demands on the self-knowledge of test takers. If this is so, formula scoring techniques would tend to be inappropriate for any multiple choice test. In this case, testers would be forced to fall back on number right scoring techniques with its accompanying decrease in score validity or rely on other, relatively untested, scoring techniques (see Frary, 1980). However, Bliss noted that there were some students who did report complying with formula scoring directions and who seemed to be able to make these distinctions. This suggests that the aforementioned ability may be a learned skill and that it may be possible to teach this skills to examinees prior to their taking tests using formula scoring directions. The purpose of this study was to test the effectiveness of such a teaching strategy.

Method

Subjects

A sample of 140 students was randomly selected from the fourth, fifth, and sixth grades of two public schools in St. Thomas, U.S. Virgin Islands and an additional 140 from two St. Thomas private schools for a total of 280 subjects. Members of each of these groups of 140 were randomly assigned to one of the four groups of a Solomon Four Group design (see Campbell and Stanley, 1963) yielding 70 subjects in each group.

Pretest directions and administration

A test composed of 36 randomly selected items from the Mathematics Concepts subtest M-1 of Form 6 of the Iowa Test of Basic Skills (Hieronymous and Linquist, 1971) was administered to two of the four student groups (designated "pretest groups"). This form is designated as being most appropriate for sixth graders and had never been used in any of the schools in which the sample students studied. Testing two thirds of the subjects above grade level was done in order to obtain larger numbers of omitted responses than might have been obtained were subjects tested at grade level.

The test directions were projected on a large screen and read aloud at the same time. The portion of those which involved the method of scoring to be used and the manner of scoring are presented below.

> Your score will be the number of questions you answer correctly minus the number you answer incorrectly. You should answer questions even when you are not sure that your answer is correct. This is especially true if you know one or more of the choices is wrong or if you have a "feeling" about which answer is correct. However, it is better to leave a blank than to guess wildly.

These directions are taken from Davis (1967) and were found to be understandable to fourth, fifth, and sixth grade students by Bliss (1980).

Examinees were directed to turn in their answer sheets and their test papers when they were finished. They were then given a reading assignment, unrelated to the test content, to do while other examinees completed the test. At no time were they allowed to look up answers to the test or allowed to discuss the test with their fellow examinees.

After all examinees had completed the pretest (never more than 35 minutes) they were asked to put away the reading assignment and their answer sheets and test apapers were returned. They were then given red pens and told that their scores would actually consist of the total number of items they marked correctly, regardless of color, and were directed to answer all the items they left blank even if this meant taking a wild guess. Examinees were permitted to work until they had completed the task.

Training procedures

One of the pretested groups and one of the nonpretested groups was chosen, on a chance basis, to receive the training strategy. This consisted of three hours of instruction conducted in three one hour sessions, each one one week apart. The first session occurred one week after pretesting was completed. The training strategy consisted of three phases, each taking approximately one hour.

Phase I consisted of explaining the process of formula scoring, the presentation of an algorithm for attacking multiple choice questions, and the presentation of the algorithm in the form of a flow chart (Figure 1). Examinees were then provided with sample multiple choice items, in increasing difficulty, to use in practicing the use of the algorithm. Their levels of success were recorded so that each student could compare earlier and later efforts (see Figure 2).

Phase II involved the use of latent image printing techniques. Students were provided with collections of multiple choice items, arranged in order of increasing difficulty, with boxes to the left of each alternative. Printed in each box was a "C" if the answer was correct or a "W" if the answer was wrong using a latent image process in which the letter was invisible until the box was colored in with a special marker. Students were directed to use the item attack algorithm to fill in the boxes of all items they believed were incorrect until they marked the box labeled "C" or until they were left with a single choice. The procedures were essentially those described by Pressy (1950). In this manner, students received immediate feedback as to their success in using the algorithm.

Phase III involved the taking of two short practice tests under formula scoring conditions. Tests were self-scored by students who were then required to fill in items they left blank. These were also self-scored and the proportions of blank items marked correctly were listed on a blackboard and compared to the expected values. This procedure of testing and self-scoring was repeated again, the results listed, and the results of this repeated testing were compared with the results of the first.

Posttesting directions and administration

All four groups were administered a 36 item test composed of questions from Form 5 of the Mathematics Concepts subtest of the Iowa Test of Basic Skills. These were analogous to the items chosen from Form 6 and used on the pretest. Identical directions and procedures were used in the pottesting situation as were used in pretesting. Posttests were administered three weeks after the last training session.

Results

Descriptive data

Descriptive statistics for the test scores of the 242 examinees who omitted one or more items are presented in Table 1. The test taking behavior variable was examined in two ways. First, the correction for guessing formula was applied to the items omitted under the first set of directions. If examinees were using partial information as directed, the expected value of these "red-formula scores" would be zero. Second, the proportion of those items prviously omitted, but then guessed correctly was determined. The expected value of these scores would be .25 since there were four choices for each item on the test. These two measures would only be linear transformations of each other if each examinee had omitted the same number of items. Table 2 presents the red-formula scores and proportion correct for each achievement test administration.

Effects of pretesting

A 2x2 analysis of variance using unweighted means techniques carried out using pretesting (group pretested or not pretested) and the presence or absence of the training procedure as the independent variables revealed no significant main effects for pretesting or pretesting by treatment interactions at the $\alpha = .05$ level for posttest proportion guessed correctly. However, there were slight ($.01 < \alpha < .05$) main effects for pretesting, but no pretesting by treatment interactions using red-formula posttest scores as the dependent variable (Tables 3 and 4). While this significant pretesting effect was disturbing, the lack of any interactions and the absence of a pretesting main effect on the other critereon measure produced the decision to continue the analysis pooling posttest scores on both variables.

Effects of training

Table 5 contains pretest and posttest red-formula and proportion guessed correctly scores. In neither of the pretested groups does the 95% confidence interval contain the expected values of either of the critereon measures. Hence, it does not appear that either of the pretested groups were using partial information in the manner perscribed by the test directions prior to training. Posttest mean scores for both variables, however, show the expected values within the 95% confidence intervals for the groups receiving the training on using partial information. The mean posttest values for red-formula and proportion guess correctly scores are significantly higher than the expected values (0 and .25, respectively) in the untrained group as indicated by the fact that the 95% confidence intervals for these means do not include the expected values.

These data indicate that during pretesting examinees, on the average, tended to omit items under formula scoring directions which they had a better than chance probability of answering correctly. On the posttest, however, the trained group appeared, on the average, to omit only those items for which they could respond with no better than a random guess. The untrained group of examinees continued to omit items for which they could have responded with more than simply a random guess.

Discuss

The training procedure, designed on the principle of presentation of an algorithm for attacking items with subsequent practice and immediate feedback, appeared to be an effective strategy for training examinees to use partial information when taking multiple choice tests. This is encouraging for a number of reasons. Lord (1975) has shown, theoretically, that when examinees use partial information in the manner described herein, formula scores are more valid indicators of true score than are number right scores. Indeed, users of testing results assume that examinees are making their best efforts when taking tests. This best effort must include making use of all information available when responding to test questions. By definition, then, examinees who are not using partial information are not making their best efforts and this should have a negative effect on the validity of obtained scores.

Second, the results of this study seem to indicate that the ability to use partial information in taking multiple choice tests seems to be a skill that can be learned and learned rather quickly and effectively. If the results of this study were to be validated over various populations, schools could include such instruction as part of their basic curricula.

This study generates a number of unresolved questions, however. First, there is the question of the stability of this training over time. Since the time between the last training session and the posttest was only three weeks, there is no indication that examinees will retain this skill over time. Second, while tests on the means of group scores showed gains in the predicted direction, the variances of scores on both critereon measures was certainly not zero. There are differences in the effects of training among examinees. Further work needs to be done to examine examinee characteristics that may interact with the training procedures. The author suggests that a good place to begin looking for these could be among those variables which tend to interact with achievement, in general, including such things as past achievement (see Bliss,1980), conceptual level (Hunt, 1970) IQ scores, and socioeconomic status.

Finally, the results of this study need to be replicated among other populations where previous studies have shown an inability of examinees to use partial information. This study can be viewed as the first of a series to investigate the effectiveness of a particular teaching strategy.

References

Bliss, L.B. A test of Lord's assumption regarding guessing behavior on multiple-choice tests using elementary school students. Journal of Educational Measurement, 1980, 17, 147-153.

Campbell, D.T. & Stanley, J.C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.

Crehan, K.D., Candor, C.A., & Beckett, G.W. Utilization of partial knowledge on objective exams under formula- and number-right scoring directions. Paper presented at the Annual Meeting of the American Educational Research Association, San Fransisco, 1976.

Cross, L.H., & Frary, R.B. An empirical test of Lord's theoretical results regarding formula-scoring of multiple choice tests. Journal of Educational Measurement, 1977, 14, 313-321.

Davis, F.B. A note on the correction for chance success. Journal of Experimental Education, 1967, 36, 42-47.

Frary, R.B. The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. Applied Psychological Measurement, 1980, 4, 79-90.

Hieronymous, A.N., & Linquist, E.F. Iowa Test of Basic Skills (Levels ed.) Forms 5 & 6, Boston: Houghton Mifflin, 1971.

Hunt, D.E. A conceptual matching model for coordinating learner characteristics with educational approaches. Interchange, 1970, 1 (2), 3-16.

Lord, F.M. Formula scoring and number-right scoring. Journal of Educational Measurement, 1975, 12, 7-11.

Poggio, J.P., Asmus, E.P., & Levy, J. An investigation of responses to omitted items under formula-scoring directions. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, 1978.

Pressy, S.L. Development and appraisal of devises providing immediate automatic scoring of objective tests and concomitant self-instruction. The Journal of Psychology, 1959, 29, 419-447.

Sheriffs, A.C., & Boomer, D.S. Who is penalized by the penalty for guessing? Journal of Educational Psychology, 1954, 45, 81-90.

Slakter, M.J., Juliano, D.B., & Sarnacki, R.E. Is there a penalty for guessing? Paper presented at the Annual Meeting of the American Educational Research Association, San Fransisco, 1976.

Table 1

Achievement Test Pre- and Posttest Statistics

Group	Pretest						Posttest					
	Formula score			Number right			Formula score			Number right		
	$\bar{X}$	SD	KR-20	$\bar{X}$	SD	KR-20	$\bar{X}$	SD	KR-20	$\bar{X}$	SD	KR-20
Pretest/Treatment (n=60)	16.83	8.10	.88	20.88	8.66	.90	20.88	7.10	.92	24.93	8.87	.89
Pretest/Control (n=68)	17.63	8.82	.90	22.20	7.03	.84	18.87	7.53	.90	23.76	6.24	.88
Nonpretest/Treatment (n=55)							20.51	7.77	.91	24.85	9.62	.89
N[illegible]est/Control ([illegible]59)							15.11	8.27	.93	20.71	6.91	.87

TABLE 2

Summary of Red-Formula and Proportion Correct Scores

	Pretest				Posttest			
	Red-Formula		Prop. Cor.		Red-Formula		Prop. Cor.	
Group	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Pretest/Treatment (n=60)	.74	1.61	.35	.29	.17	1.11	.26	.21
Pretest/Control (n=68)	1.09	1.78	.37	.27	1.21	1.81	.39	.27
Nonpretest/Treatment (n=55)					.05	.92	.28	.15
Nonpretest/Control (n=59)					.56	1.56	.33	.30

Table 3

Analysis of Variance Source

Table for Rod-Formula Scores

Source	SS	df	MS	F
Pretesting	9.758	1	9.758	4.93*
Training	34.972	1	34.972	17.67*
Pretesting X Training	4.728	1	4.728	2.38
Within cells	471.039	238	1.979	

* $p<.05$

Table 4

Analysis of Variance Source

Table for Proportion Guessed Correctly

Source	SS	df	MS	F
Pretesting	.061	1	.061	1.15
Training	.303	1	.303	5.71 *
Pretesting X Training	.182	1	.182	3.43
Within cells	12.518	238	.053	

* $p < .05$

Table 5

Pooled Pretest and Posttest Red-Formula and Proportion Correct Scores

	N	$\bar{X}$	95% confidence interval
Pretest			
Red-Formula Scores			
Treatment Group	60	.742	.326 to 1.163
Nontreatment Group	68	1.090	.663 to 1.512
Proportion Guessed Correctly			
Treatment Group	60	.351	.277 to .425
Nontreatment Group	68	.365	.301 to .429
Posttest			
Red-Formula Scores			
Treatment Group	115	.114	-.076 to .304
Nontreatment Group	127	.910	.618 to 1.202
Proportion Guessed Correctly			
Treatment Group	115	.275	.241 to .309
Nontreatment Group	127	.361	.307 to .414

Figure 1

Flow Chart of the Item Attack Algorithm

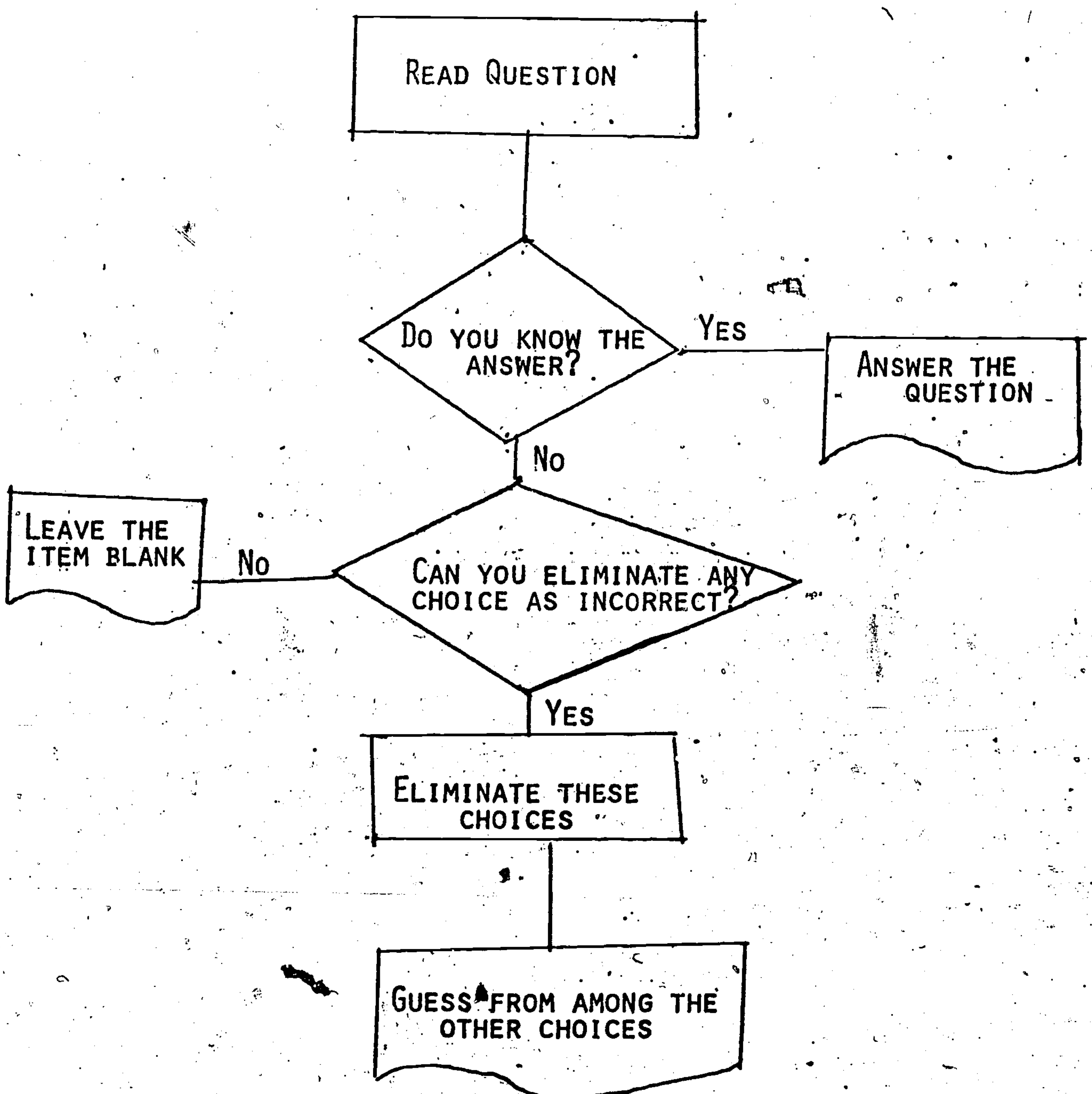

Figure 2

Model Phase I Practice Page

THE CAPITAL CITY OF IOWA IS

A. ALBANY

B. DES MOINES

C. SOUIX CITY

D. WASHINGTON, D.C.

1. DO YOU KNOW THE ANSWER?

   A. YES - WRITE THE LETTER OF THE ANSWER HERE. ________

   B. NO - GO TO STEP #2

2. CAN YOU ELIMINATE ANY OF THE CHOICES AS BEING WRONG?

   A. YES - LIST THE LETTERS OF THE WRONG ANSWERS HERE.

   ______ ______ ______

   NOW, TAKE A GUESS FROM AMONG THE CHOICES YOU DON'T KNOW ARE WRONG AND LIST THE ANSWER HERE. ________

   B. NO - GO ON TO THE NEXT QUESTION.