DOCUMENT RESUME

ED 228 281 TM 830 158

AUTHOR Shaha, Steven H.
TITLE Reduced Anxiety and Increased Assessment Effectiveness with Matching Test Formats.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO CSE-R-192
PUB DATE 82
NOTE 22p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Multiple Choice Tests; *Objective Tests; Response Style (Tests); Secondary Education; Self Evaluation (Individuals); Student Reaction; *Test Anxiety; *Test Format; *Testing Problems

ABSTRACT

Traditionally, matching test formats have been avoided in favor of multiple-choice items for several reasons, including item analysis properties and chance performance characteristics. In the light of research on test format and anxiety, this study postulates that, if a matching test could assess knowledge for a given topic as effectively as an analogous multiple-choice test, yet present a less threatening, less anxiety-provoking situation, then the matching format should be utilized for testing. Two experiments measured the comparative effectiveness of each format for assessing student recall capabilities and for reducing test anxiety. Sixty-four students from Los Angeles area high schools were first administered 12 premise/response matching pair items and 12 multiple-choice items to assess prior knowledge. Test anxiety was measures with an experimental Test Anxiety Inventory, and test preference was measured with a three-item questionnaire. The second experiment determined generalizability, testing recall of novel or recently encoded material. Test subjects consistently favored the matching tests, scored equally high on them, and experienced significantly less debilitating test anxiety. The reduction of anxiety was possibly due to successful test-taking strategies and positive self-evaluations during testing. (CM)

REDUCED ANXIETY AND INCREASED ASSESSMENT EFFECTIVENESS WITH MATCHING TEST FORMATS

Steven H. Shaha

CSE Report No. 192
1982

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

The research reported herein was supported in whole or in part by a grant to the Center for the Study of Evaluation from the National Institute of Education, U. S. Department of Education. However, the opinions and findings expressed here do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

TABLE OF CONTENTS

INTRODUCTION

Research since the early 1950's has consistently verified the significantly negative relationship between test anxiety and academic achievement (Shaha & Wittrock, Note 1). S. B. Sarason and Mandler (1952) were among the first to uncover a significant correlation between test scores and test anxiety. In similar research, Alpert and Haber (1960) found that both grade point averages (GPA) and examination scores are predicted by test anxiety, and I. G. Sarason (1963) showed that standardized test scores in mathematics and verbal skills are also predicted by test anxiety.

Efforts to define the antecedent causes of test anxiety have led to various interpretational theories. Nicholls (1976) defined test anxiety as "self-evaluation," stating that test anxiety scores actually reflect students' perceptions of their own inadequacies in testing situations. Gaudry (1977) supported a similar theory, proposing that test anxiety is caused by previous failures in testing situations. Hill (1972) and Kirkland and Hollandsworth (1980) have proposed that test anxiety is caused by poor test-taking skills. They independently concluded that highly anxious childrens' lower test scores and lower school achievement stems from inadequate test-taking strategies rather than from learning deficiencies. High anxiety coupled with poor test-taking skills interfere with the effective completion of tests.

Several treatments have been designed to reduce test anxiety in an effort to increase academic achievement. Golfried, Linehan, and Smith's (1978) use of cognitive restructuring techniques reduced test anxiety and raised test scores. Similarly, Miechenbaum (1972) increased test scores through cognitive modification techniques which familiarized subjects with their anxieties and then offered them systematic desensitization treatments or ideal models to follow. Williams and Hill (Note 2) reduced test anxiety and increased test scores of high anxiety students by modifying test instructions. Changing instructions so that the testing situation appeared less evaluative and threatening increased the subjects' scores significantly. The altered instructions, however, caused a decrease in test scores for middle and low anxiety subjects.

The critical issue remains whether or not test anxiety can be effectively reduced, and test scores subsequently raised, without any negative effects such as lowered scores for normally or low anxious students and without resorting to costly treatments or special programs. The question arises as to whether altering the <u>form</u> of a test, and not merely the instructions, would have these desired effects on anxiety. In short, is there a less threatening format for which most students have effective test-taking strategies and which will efficiently assess students' knowledge of a given subject area?

An informal questionnaire was administered to 150 students between the ages of 8 and 26. The inquiry asked for free responses to only one question: "Which type of test or test question makes you worry the least?"

Responses included oral exams, essays, fill-in-the blank completion questions, and others. One predominant response, however, was matching tests. Traditionally, matching test formats have been avoided in favor of multiple-choice items for various reasons, including item analysis properties and chance performance characteristics (Popham, 1981; Shaha, Note 3). Howeve if a matching test could assess knowledge for a given topic area as effectively as an analogous multiple-choice test, yet present a significantly less threatening, less anxiety-provoking situation, then the matching format should be utilized instead of the traditional alternative. It was upon this logic that the following research was conducted.

Two experiments were designed to measure the comparative effectiveness of analogous multiple-choice and matching tests for (1) assessing student recall capabilities, and (2) for reducing test anxiety. It was anticipated that the matching test format would represent a significantly less anxiety-producing stimulus and yet be equally effective for measuring subject recall. Measurement effectiveness was determined to be representable by item discrimination and difficulty.

EXPERIMENT I: METHOD

Subjects and Design

Sixty-four juniors and seniors from West Los Angeles area high schools participated in three classroom groups (19, 22, 23 students) as voluntary subjects.

Materials and Tasks

Twelve premise/response pairs were composed dealing with facts (premises) about past Presidents of the United States (responses). All pairs related to one common stem: "Which of the following Presidents listed ... would you associate with the statement(s) given?" A matching test was constructed using the 12 test pairs. Premises were listed vertically on the left side of the test sheet, and responses on the right side. Three extra Presidents' names were added to the response list as distractors. Premises were numbered and responses lettered, and all were randomly ordered. A blank space was provided next to each premise for recording the letter corresponding to the selected response.

Twelve multiple-choice test items, drawing upon the response alternatives and the additional three distractors described above, were constructed. Each item had the same basic stem, one of each of the 12 premises as the questions, and four alternative response choices. Each of the 12 responses was used as an alternative three times, and the three additional distractors were used four times each. The completed test was presented in a three-page booklet.

Test anxiety was assessed via a posttest questionnaire based on the Inventory of Test Anxiety (Osterhouse, 1972). The resulting experimental measure, hereafter referred to as the Test Anxiety Inventory (TAI), consisted of 16 likert-scaled items and was designed to measure anxiety felt during the test as reported in retrospect.

Test preference was assessed by a three-item questionniare with the following questions: (1) "Which of the two test forms did you prefer?" (2) "Which test was easiest for you?", and (3) "Which test was faster?"

All materials and tasks described above were reproduced by photocopying on 8½ x 11 inch (21 x 27½ cm) standard sized paper. Each of the tasks was stapled into separate booklets preceded by an instruction sheet.

Procedures

Subjects participated in their regular classroom groups in a design counterbalanced for test sequence. Tests were distributed from a randomly shuffled stack consisting of half of each test type. After subjects read instructions for their tests silently and all procedural questions were answered individually, all subjects initiated the test tasks simultaneously. Upon completion of the first test, an experimenter equipped with a stopwatch collected the test, recorded the time-on-task (to the nearest 30 seconds), and gave the student a TAI for completion. After completing the TAI, the subject was distributed the second, opposing test form. Completion of the second test was also followed by a second TAI. The final subject task was to respond to the test preference questionnaire, after which students were dismissed from the room.

No time limits were imposed for any experimental tasks. However, as mentioned, time-on-task was monitored by recording starting and finishing times for each of the two tests.

Results and Discussion

A one-way analysis of variance for time-on-task revealed no differences between groups. Both test forms were scored for number of correct responses. A two-way analysis of variance for test scores x test sequence revealed no significant differences between scores for either test format, and no significant main effect for test sequence counterbalancing. There was also no significant interaction.

- - - - - - - - - -

Insert Table 1 here

- - - - - - - - - -

In order to compare assessment effectiveness between the traditional multiple-choice format and the matching test alternative, detailed analysis of item difficulty ratings and discrimination properties were performed. Item discrimination refers to the consistency with which high scoring subjects respond correctly to an item while low scoring students err, meaning that the test truly discriminates between those who possess and those who do not possess requisite knowledge.

An item analysis was completed for each test. Item difficulty was calculated as the proportion of responding students who scored incorrectly (high proportion=high difficulty). Item discrimination was calculated as the point-biserial correlation of correct/incorrect response patterns

Table 1

Means and Standard Deviations:

Experiment I

		Matching Test	Multiple-Choice Test
Time-on-Task	Mean	5.68	5.32
	SD	1.43	.99
Number of Correct Responses	Mean	9.62	9.38
	SD	2.53	3.10
Item Difficulty	Mean	.32	.38
	SD	.18	.21
Item Discrimination	Mean	.73**	.53
	SD	.23	.11
Test Anxiety	Mean	2.33	4.07**
	SD	.49	1.12

** Significantly greater at $p=.01$

with total test scores. Analysis of variance for item difficulty yielded no significant differences between groups, meaning that neither test was significantly harder or easier for the students. Analysis of variance for item discrimination, on the other hand, yielded a significant F ratio in favor of the matching test ($E(1,62)=9.41$, MSerr = .21, $p<.01$). In other words, the matching test more accurately discriminated between high and low scorers (see Table 1), where high scoring subjects were more consistently correct on the matching test.

The Likert-scale responses to test anxiety questionnaires were reduced to a mean anxiety score for each subject, on each test. Analysis of variance for anxiety showed that the matching test produced significantly less test anxiety than the multiple-choice test (see Table 1). The implication is that the matching test format presented a significantly less threatening situation and hence produced significantly less test anxiety. The test preference questionnaires, in which 63% of the subjects stated that they preferred the matching test format, supported the conclusion that the matching test is less threatening. The majority of the respondents also claimed the matching test was both easier (83%) and took less time to complete (53%). The claims by subjects concerning the comparative time taken to complete the tests were especially interesting in view of the fact that no significant differences were found for actual time-on-task. This phenomenon was previously discovered and discussed by I. G. Sarason and Stoops (1978).

Considered as a whole, the results of Experiment I support the assertion that matching tests offer a significantly less anxiety-producing

format, as evidenced by anxiety scores and preference reports. Further, as indicated by the significantly higher item discrimination indices for the matching test items, the reduction in anxiety does not reduce test effectiveness for discriminating between subjects familiar with topic material and those with less knowledge.

EXPERIMENT II: METHOD

The tests in Experiment I were designed to assess subjects' ability to respond to questions based on prior knowledge. A second experiment was conducted to determine whether the results of the first experiment were generalizable to tests covering material either novel to or just encoded by the subject.

Subjects and Design

The same 64 high school juniors and seniors from Los Angeles area schools participated in the identical classroom groups one week later.

Materials and Tasks

Following the same procedures used in the first experiment, tests covering information about two topics were constructed: (1) Whales, and (2) Far Eastern Religions. Twelve premise/response pairs were developed for each topic and then converted into analogous matching and multiple-choice tests. For encoding purposes, prose passages were then composed based on the questions, and the passages were taped on cassettes.

Test anxiety, test format preference, and time-on-task were all measured by devices identical to those used in the first experiment.

Procedures

Data were collected on separate, consecutive days for each of the two topics. As in Experiment I, subjects completed the experimental tasks in the following sequence: (1) Test format #1 (format determined by random distribution procedures), (2) TAI #1, (3) Test format #2, (4) TAI #2, and (5) Test preference questionnaire.

On the first day, subjects listened to the taped passage about Whales (3 min. duration) while they read the identical passage silently. This procedure was designed to maximize encoding. Instructions for the encoding task warned subjects that they would be tested for their memory of the stimulus information, but no reference was made to the mode or manner of testing. The remaining experimental tasks were performed without any further exposure to the stimulus material. The same procedure was employed on the second day with the tape (3.5 min. duration) and passage about Far Eastern Religions.

Results and Discussion

Scoring procedures were identical to those employed in the first experiment. Analysis of variance for time-on-task yielded no significant difference for either test format, despite topic matter (see Table 2). A two-way analysis of variance for each topic area yielded no significant effects for number of correct responses, for test sequence effects (counterbalancing), or for the interaction.

Item analysis were conducted for all four tests. The tests measuring recall of Whales revealed no significant differences for item diffi-

culty or for item discrimination statistics. The test covering Far Eastern religions showed no significant differences for item difficulty, but did produce a significant F ratio by analysis of variance between test types for time discrimination ($F(1,62)=5.63$; MSerr=.13; $P<.05$). It appears that the matching tests were at least as effective assessment tools as the multiple-choice formats.

- - - - - - - - - - -

Insert Table 2 here

- - - - - - - - - - -

Test anxiety data from both topics assessed mirrored results from Experiment I. Matching test formats were significantly less anxiety-producing for both Whales ($F(1,62)=6.21$; MSerr=.17) and Far Eastern Religions topics ($F(1,62)=10.03$; MSerr=.33). Test preference was also decidedly in favor of the matching formats. Actual percentages of subjects stating a preference for the matching test mode were 79% for Whales and an overwhelming 93% for Far Eastern Religions. Questionnaires again consistently echoed the findings that subjects perceived a shorter time-on-task for the matching test formats (68%, 54% respectively), even though analyses for time-on-task revealed no significant differences between formats. Subjects also rated the matching test as easier (73%, 91%), while no significant differences for scores were found.

CONCLUSIONS

The two experiments considered together clearly support the use of matching test formats for assessing either prior knowledge or recall of recently encoded material. Although test developers and theorists may debate use of the matching test format (Shaha, Note 3; Burry, 1971)

Table 2

Means and Standard Deviations:

Experiment II

Topic		Whales		Far East Religions	
Task		Matching Test	Multiple-Choice Test	Matching Test	Multiple-Choice Test
Time-on-Task	Mean	4.85	4.96	7.34	6.86
	SD	.79	.55	2.39	1.88
Number of Correct Responses	Mean	10.03	9.85	7.38	6.33
	SD	3.11	3.69	4.21	3.28
Item Difficulty	Mean	.45	.49	.67	.65
	SD	.42	.27	.41	.28
Item Discrimination	Mean	.63*	.44	.78*	.68
	SD	.11	.32	.14	.12
Test Anxiety	Mean	2.06	3.81**	2.89	4.42**
	SD	1.31	.62	2.03	41

* Significantly at $p=.05$
** Significantly at $p=.01$

these experiments suggest that multiple-choice tests should not necessarily be preferred for either assessment effectiveness or anxiety reduction when contrasted with the matching format. On the contrary, subjects overwhelmingly and consistently favored the use of matching tests, scored equally high on them, and experienced significantly less debilitating test anxiety.

Perhaps the most interesting finding in these studies involves the reduction of test anxiety, without any apparent ill effects, merely by changing test format. This finding cannot be overemphasized. The correlation between anxiety and both test performance and scholastic achievement in general raises major concerns about the use of any assessment technique which might unnecessarily increase anxiety and decrease test performance.

One possible explanation for the reduction in test anxiety discovered in these studies lies in successful test-taking strategies and positive self-evaluations. Shaha (Note 3) found that subjects employ simple elimination strategies when responding to matching test items; the easier matches are made first, and made quickly and easily. The subject is immediately reinforced, and his/her confidence increases as the elimination strategies are found to be successful. Although the simple matches are expended as the student proceeds and encounters more difficult associations, the initial optimism does not wear-off, as is evidenced by post-experimental test preferences and post-test anxiety scores.

In summary, the "self-evaluation" theories discussed earlier are

supported here (e.g., Gaudry, 1977; Nicholls, 1976) as well as the "test-taking skill" proposals (e.g., Hill, 1972; Kirkland, 1980). On the basis of their effective elimination strategies, students feel reduced anxiety and increased competence. However, the optimism and subsequent reduced anxiety are a student perception. Test scores, time-on-task, and item difficulty data discount any actual superiority of the matching test for ease or efficiency of strategy.

Reduction of test anxiety cannot be overemphasized. Since test anxiety predicts both test scores and scholastic achievement in general, any assessment technique which might unnecessarily increase anxiety should be avoided. If a particular test format can lower anxiety and yield outcome scores and assessment data equivalent to those obtained with other formats, then the anxiety-reducing method should be employed. Certainly further research by test developers is in order.

REFERENCE NOTES

1. Shaha, S. H., & Wittrock, M. C. Cognitive processes related to school achievement. Unpublished manuscript, University of California, Los Angeles, 1982.

2. Williams, J. P., & Hill, K. T. Performance on achievement test problems as a function of optimizing test presentation instructions and test anxiety. Unpublished manuscript, University of Illinois, Urbana-Champaign, 1976.

3. Shaha, S. H. Response strategies revealed through item analyses of matching tests. Unpublished manuscript, University of California, Los Angeles, 1982.

REFERENCES

Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Burry, J. Constructing test items to measure educational objectives. Center for the Study of Evaluation: University of California, Los Angeles, 1971.

Gaudry, E. Studies of the effects of experimentally induced experiences of success or failure. In C. D. Speilberger & I. G. Sarason (Eds.), Stress and anxiety. Washington, D. C.: Hemisphere Publication Corporation, 1977.

Golfried, M., Linehan, M., & Smith, J. Reduction of test anxiety through cognitive restructuring. Journal of Consulting and Clinical Psychology, 1978, 46, 32-38.

Hill, K. T. Anxiety in an evaluative context. In W. W. Hartup (Ed.), The young child (vol. 2). Washington, D. C.: National Association for the Education of Young Children, 1972.

Kirkland, K., & Hollandsworth, J. G., Jr. Effective test-taking: Skills-acquisition versus anxiety-reduction techniques. Journal of Consulting and Clinical Psychology, 1980, 48, 431-439.

Miechenbaum, D. H. Cognitive modification of test anxious college students. Journal of Consulting and Clinical Psychology, 1972, 39, 370-380.

Nicholls, J. G. When a scale measures more than its name denotes: Test anxiety scales for children. Journal of Consulting and Clinical Psychology, 1976, 44, 976.

Popham, W. J. Modern educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1981.

Sarason, I. G. Critique and notes: Test anxiety and intellectual performance. Journal of Abnormal and Social Psychology, 1963, 66, 73-75.

Sarason, I. G., & Stoops, R. Test anxiety and the passage of time. Journal of Consulting and Clinical Psychology, 1978, 46, 102-109.

Sarason, S. B., & Mandler, G. Some correlates of test anxiety. Journal of Abnormal and Social Psychology, 1952, 47, 810-817.