DOCUMENT RESUME

ED 212 672 TM 820 118

AUTHOR Slem, Charles M.
TITLE Modifying the Multiple Choice Test: A Short Answer Option.
PUB DATE 25 Aug 81
NOTE 21p.; Paper presented at the Annual Convention of the American Psychological Association (Los Angeles, CA, August 25, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Constructed Response; *Essay Tests; Higher Education; *Multiple Choice Tests; Response Style (Tests); *Test Construction; *Test Format; Testing Problems

ABSTRACT

Over the years many criticisms have been offered against the multiple choice test format. Ambiguous, and emphasizing isolated information, they are also the most difficult objective tests to construct. Over-interpretation is a danger of multiple choice examinations with students picking subtle answers the test makers consider incorrect. Yet, the multiple choice test is a tool offering versatility in measuring educational objectives. It measurers the student's discriminatory thinking, comprehension, application, synthesis, and evaluation. Guessing is minimized and the greater number of items on the exam, the more representative are the sample questions. Perceived ambiguity could allow the scorer to see that the student understood the concept being tested by explaining particular qualifications they perceive and their individual resolutions of the ambiguity. The multiple choice-essay exam (MCE) was developed incorporating a wide right-hand margin so the student could explain alternate answers. This document deals with its development and implementation. (Author/CE)

MODIFYING THE MULTIPLE CHOICE TEST: A SHORT ANSWER OPTION

Charles M. Slem, Ph.D.
Psychology Department
California Polytechnic State University

1981 Annual Convention of the
American Psychological Association
Los Angeles, California
August 25, 1981

Milton and Edgerly (1976), McGuire (1968), Green (1975), Black (1963), and Macintosh (1974) summarize the many criticisms of the multiple choice test format. Students also offer complaints that multiple choice test items tend to be ambiguous and often emphasize the trivial.

McGuire (1968) reports that over half of the multiple choice items he studied stressed recall and recognition of isolated information. Green (1975) points out that multiple choice tests are the most difficult objective tests to construct. For example, it is often difficult to construct a set of alternatives to a question with only one alternative as clearly justifiable; and it is relatively easy to omit one or more of the qualifications needed for correct answer selection. In addition, there is a tendency for test makers to include nonfunctional words which interfere with student comprehension. Difficulty level is also hard to control, and Black (1963) reports that as a result multiple choice tests might be detrimental to brighter students. According to Black, multiple choice tests are apt to contain questions which might confuse brighter students who pick subtle answers which the test maker considers incorrect. Millman and Pauk (1969) also warn of the danger of overinterpreting multiple choice questions.

Yet, multiple choice tests are quite common in college settings and represent the most popular objective testing technique (Nunnally, 1972). Green (1975) maintains that the multiple choice test is the best type of objective test for measuring a variety of educational objectives. Multiple choice tests are versatile and require discriminatory thinking on the part of the student. The test format is useful for measuring knowledge,

comprehension, application, synthesis, and evaluation. It minimizes guessing found in true-false tests and is adaptable to numerous fields.

Multiple choice tests have several other inherent advantages. The greater the number of items on an exam, the more representative are the sampled questions. Multiple choice tests are also more objective than other formats, e.g., essay exams, because two or more scorers will arrive at the same score after a key is prepared. According to Milton and Edgerly (1976), however, unless the test is constructed carefully, much of the score could be arbitrary.

Attempts have been made to modify the multiple choice format so that its strengths could be retained while alleviating several of its major criticisms. The most frequently cited modification is the partial knowledge or confidence-testing procedures. (Lumsden, 1977, Echternacht, 1972, Curlette, 1978, Coombs, et. al., 1956, Stanley and Wang, 1970, Diamond, 1975) One variation (Jacobs, 1971) has the student assign personal probabilities of correctness to each alternative, and the student's personal probability of the actual correct alternative becomes his/her score for the question. Another system (Coombs, et. al., 1956) scores questions by having the students eliminate the alternatives they decide are incorrect. The more alternatives appropriately marked incorrect, the more credit the student earns.

Although confidence-testing allows credit to be assigned for partial knowledge, it is not without its criticisms. Curlette (1978) found that there is a method of responding that will increase the student's score irrespective of their knowledge. On a practical level, confidence testing formats seem complicated and time consuming to score. A more basic criticism, however, is that the confidence-testing option does not squarely address the issues of ambiguity, overinterpretation, and confusion. The literature

review revealed it is difficult and time consuming for an instructor to create a test which is perceived to be error free and not subject to multiple interpretations which may have inadvertantly been given in lecture.

One strategy for correcting the distraction of perceived ambiguity, confusion, and lack of all necessary qualifications would be to allow students an alternative method of expressing themselves on those items. This mode would serve to indicate to the scorer that the students understood the concept being tested by explaining the particular qualifications they perceive and their individual resolutions of the ambiguity and confusion. Argent (1974) found that students are good judges of ambiguity and can easily point out why an incorrect alternative could be an appropriate response. This strategy led to the development of a multiple choice exam with a short essay option which came to be known as the multiple choice-essay exam (hereafter referred to as "MCE"). Structurally, the test booklet appears to be a multiple choice test. The only apparent difference is an extremely wide right-hand margin, 3 1/2" to 4" wide. (See Attachment 1) The answer sheet is a standard machine scored form, in this case Scantron #882.

The student was instructed that the test items could be answered in two ways. Each question could be answered as if it was a multiple choice question. If, however, the student found that the question contained any of the following problems--ambiguity, the appearance of no right answers, the appearance of more than one right answer, different right answers depending upon how the question is interpreted--the student was instructed to place on the machine scored answer sheet the answer he or she felt was the one the instructor was most likely to consider correct. Then the student was to write in the wide

margin next to the question a short essay explaining the answer they believed to be correct and why.

When the machine scored forms were returned to the instructor, the short-answer essays were reviewed. If the student had gotten the question correct, the short-answer essay was not read. If the student had gotten the question wrong, the short-answer essay was carefully read and from zero to full credit was assigned.

In assessing the effectiveness of this test format, several outcomes were considered important.

1. Did the students find this format to be helpful?

2. How did the MCE format compare with the exam format that students usually prefer?

3. Did this option significantly raise scores on examinations?

4. How often were essays written and how much time was necessary to score them?

5. Was there any relationship between the number of essays written and item difficulty?

Method

At the beginning of the class meeting following the exam, an anonymous questionnaire (see Attachment 2) was administered to 118 students in 3 classes. The questionnaire included forced-choice items regarding the perceived difficulty of the MCE exam, its compairson with students' preferred exam formats, as well as its compairson with the standard multiple choice format. Other questions surveyed students' preferred test format, reasons for their preference, and the

most frequently encountered test format. Finally, students were asked how many essays they had written and whether or not the MCE format was helpful if they had or had not made use of the essay option.

Seven hundred seventy-nine exams were assessed to determine the number of essays per exam and credit given. An item analysis of 110 MCE exams was correlated with the number of essays written on each test item. MCE exams from twelve classes (N = 493) were analyzed to determine the class average before essays were reviewed, the class average after the effects of the essays were included, and the difference between the two averages. The total number of essays written was categorized by credit given for each essay (0, 1/4, 1/2, 3/4, 1, or already correct). Finally, the time to review the essay portion of the exam was calculated for seven classes in which the MCE exam was given.

Results--Questionnaire Analysis

Table 1 summarizes the students' preferred exam formats as well as the formats they most frequently encounter. In responding to the question concerning preferred test formats, a large percentage of the students, 62.6%, preferred exams with a multiple choice format. Open-ended responses revealed that the largest group of these students felt the multiple-choice test was the easiest for them to take. According to many of these students, recognition was easier than recall; and they knew the right answer was among the alternatives.

A second large group reported that their preference was due to their past success with this format. They did not amplify. Other reasons cited included a better opportunity to guess, less need for memorization, and a feeling that multiple-choice questions helped refresh their memories. Several

students were concerned with poor grammar and spelling skills, and the multiple choice test did not require the demonstration of grammar skills.

The next most preferred exam format was short essays, which 16.5% preferred. The most frequently cited reason for preferring the short-answer essay format involved criticism of the multiple-choice format. These students believed that multiple-choice questions were easy to misinterpret, and they found it often difficult to discriminate between alternatives. Short-answer essays were straightforward and allowed the students to more fully explain themselves in their own words. Several students felt this format was the best test of knowledge. An additional small number of students appreciated the opportunity to express their own views and opinions.

Only 2.6% of the total sample selected long essays as their preferred test format.

In response to the item concerning the most frequently encountered test format, 58.0% reported they encountered the multiple choice format most frequently. Combinations of exam formats accounted for 13.4% and short essays were most frequently encountered by 11.8% of the students. Over 15% reported their most frequently encountered test was other than the alternatives listed on the questionnaire. Inspection of the raw data revealed that the majority of those students had exams with a problem solving format typical of engineering and math curricula.

Of the students who preferred multiple choice tests, 62.2% actually encountered multiple choice tests most frequently. Only 15% of the students who preferred the short essay exams encountered short essay exams most frequently. The most frequently encountered exam for those who preferred short essays was the multiple choice format.

In rating the MCE format against the standard multiple choice format, 84.5% rated the MCE much better or somewhat better. (See Table 2) Only 5.2% rated it somewhat worse than the standard multiple choice format. Of those who preferred short-answer formats, 80% found the MCE format preferable to most multiple choice formats.

In rating the MCE with their preferred test format, 68.7% rated it much better or somewhat better than their preferred format. Only 9.6% rated the MCE worse or somewhat worse than their preferred test format. Of those who preferred multiple choice formats, 77.8% rated the MCE much better than multiple choice exams. Of the students preferring short-answer formats, 60% found the MCE format more helpful. Only those students who preferred long essays rated the MCE more consistently below their preferred format.

At least one essay was reported written by 41.2% of the students. Of the students writing essays, 98% felt the option of writing essays made the test easier for them. The majority of the open ended responses to this question indicated that the students appreciated the opportunity to explain or clarify their answers. Some students pointed to clarifying confusing questions, while other students clarified their answers because they themselves were confused. Several students stated that they appreciated being treated as individuals with this format. They reported they felt they could have different perceptions than the instructor and also be correct. Another small number of students reported that the option helped them organize their thoughts and enabled them to select the correct answer. A few students stated specifically that it reduced anxiety for them.

Students reporting writing no essays made up 58.8% of the sample. However, 75.4% of these students felt the option was still helpful. Their

open ended responses fell into three overlapping categories plus a residual. The most frequently cited reason seemed to center around a reduction of anxiety or tension. This reduction occurred while studying for the test and/or while taking the exam itself. The second most frequently cited reason was an appreciation that if they had encountered an ambiguous question they had the opportunity to explain their answers. The third category focused on having been given the choice to use the option if they wished.

Test Analysis

Seven hundred seventy-nine tests were analyzed; and of the 779 students taking the exam, 425 or 54.5% wrote a total of 1791 essays. Of the essays written, 64.7% were already scored as correct on the machine-scored answer sheet and were therefore not read. A total of 19.8% of the essays received partial credit (.25 to .75 points) and 5.6% received full credit. No credit was given to 9.9% of the essays.

The essay option had very little impact on the final average of the tests. The average percent change was only 0.5%, and only an average of 1.25 students per test experienced a gain which led to a raise in the letter grade for the exam. (See Table 3) The relationship between the item analysis and number of essays written produced a modest correlation of $r = .35$, $p < .003$.

The exams required an average of 23.1 minutes to score. This is an overestimate since it includes time which was used to categorize responses for research purposes. Another indication of the investment of time required was the average number of short-answer essays that had to be read per test.

The average number of essays per class per test was 64. Since an average of over 41 of those 64 had already been scored as correct on the machine scored answer sheet, an average of only 23 questions per class (class size approximately 40) required reading.

Discussion

The results of the questionnaire indicate that students believe that the MCE option of writing short answers to ambiguous or confusing multiple choice questions was definitely helpful. Even students who wrote no essays found the option anxiety reducing. Students generally preferred the MCE format over the standard multiple choice format; and in a surprising 68.7% of the sample, students rate the MCE somewhat or much better than their preferred test format. Although the evaluation of the MCE test was only conducted immediately after the examination and before the students received their grades, Attkisson (1975) found that the student evaluation of an examination immediately after the testing period and an evaluation conducted one week later (following a review of the exam, receipt of scores, and assignment of grades) were the same. In addition these results of the current study are not likely to be explained as a simple version of the Hawthorne effect because in subsequent quarters student evaluations have remained high even when the essay option was built into the course structure and given very little attention.

One reason the MCE format may be perceived as being so helpful is that it combines major elements of the two most preferred exams, multiple choice and short-answer essay. In fact, the MCE removed many of the criticisms of multiple choice tests offered by those who preferred short-answer essays. The MCE appears to be superior to each of them individually since a large percentage of those students who preferred short answer or multiple choice

rated the MCE as better than their preferred format, 60.0% and 77.8 respectively.

Another possible reason for the success of this format from the students perspective was articulated by several students who believed that this option gave them more control over the situation since their possible idiosyncratic interpretations would be considered. This option was therefore anxiety reducing regardless of whether or not a student used it. This interpretation is more consistent with a higher order version of the Hawthrone effect which suggests that one component of increased productivity is a sense of control in the work group.

It is interesting to note that although the students feel the MCE test format is helpful it does not significantly affect their scores. It only raised the class average .22 of a point, and only an average of 1.25 students received gains that raised their letter grades for the exam. The MCE, while making an insignificant impact on the class average, apparently has a great impact on the individual student's perception of the exam and its fairness.

Most of the essays were written on questions that were already correct on the machine scored answer sheet. This finding is consistent with research that demonstrates that people can make accurate responses or discriminations although they have very little or no confidence in their accuracy (Adams, 1957). It is possible that clarifying these answers may have had the effect of reducing anxiety in general and as a result may have raised the class average in general. Since it was not feasible in the study to have a control group which took a standard multiple choice exam without the MCE option, no comment can be made.

Further research will focus on control group comparisons, an indepth analysis of individual styles of responding, and the criteria by which partial credit is given.

It is obvious, however, that the MCE format is a marked improvement over the standard multiple choice format. The exam is as easy to administer as the standard multiple choice test and requires only a modest investment of additional time by the instructor. As university budgets become increasingly strained and class sizes increase, instructors who prefer essay type exams will not be able to use that format. The MCE format appears to be a viable alternative that is perceived to be fair and equitable by the students.

Table 1

Format	Preferred Exams N	Preferred Exams %	Most Frequently Encountered Exams N	Most Frequently Encountered Exams %
Multiple Choice	72	62.6%	69	58.0%
Short Essay	19	16.5	14	11.8
Long Essay	3	2.6	2	1.7
True/False	9	7.8	0	0.0
Matching	8	7.0	0	0.0
Oral	4	3.5	0	0.0
Combinations	0	0.0	16	13.4
Other	0	0.0	18	15.1
	115	100.0	115	100.0

Table 2

	MCE Compared to Standard Multiple Choice		MCE Compared to Preferred Test Format	
	N	%	N	%
Much Better	72	62.1	41	35.7
Somewhat Better	26	22.4	38	33.0
About the Same	12	10.3	25	21.7
Somewhat Worse	6	5.2	7	6.1
Much Worse	0	0.0	4	3.5
	116	100.0	115	100.0

Table 3

Number of Students	Number of Essays Written	Class Average Without Essay	Class Average With Essay	Change	Percent Change	Number of Letter Grades Changed
41	13	37.22	37.24	.02 pts.	.1%	0
31	40	39.58	39.89	.31	.1	1
41	52	39.95	40.12	.17	.4	2
32	19	41.22	41.34	.12	.3	1
33	17	40.06	40.15	.09	.2	1
36	35	39.50	39.66	.16	.4	1
27	93	41.37	42.05	.68	1.6	2
43	81	38.77	39.08	.31	.8	3
35	67	35.86	36.24	.38	1.1	0
60	46	36.57	36.78	.21	.6	2
59	44	40.57	40.71	.14	.3	0
55	49	41.42	40.52	.10	.2	2
				ave. = .22	.5%	1.25

Bibliography

Adams, J. Laboratory Studies of Behavior Without Awareness. Psychological Bulletin, 1957, 54:383-405.

Argent, B. Short Answer Questions and Objective Items. In Macintosh, H. G. (Ed.) Techniques and Problems of Assessment. London: Edward Arnold, 1974.

Attkisson, C. C., Snyder, C. R. Student Evaluation of Multiple Choice and Word Association Exams. Journal of Instructional Psychology 2(1):9-15, 1975.

Black, H. They Shall Not Pass. New York: William Morrow and Co., 1963.

Coombs, C. H., Milholland, J. E., and Warner, F. "The Assessment of Partial Knowledge," Educational and Psychological Measurement, 16:13-37, 1956.

Curlette, W. Demonstration of Response Strategies in a Confidence-Testing Procedure. Psychological Reports, 43:479-85, 1978.

Diamond, T. A Preliminary Study of the Reliability and Validity of a Scoring Procedure Based on Confidence and Partial Information. Journal of Educational Measurement, 12:129-33, 1975.

Echternach, G. J. "The Use of Confidence Testing in Objective Tests," Review of Educational Research, 42:217-236, 1972.

Green, J. A. Teacher-Made Tests, 2nd Edition. San Francisco: Harper and Row, 1975.

Jacobs, S. S. "Correlates of Unwarrented Confidence in Responses to Objective Test Items," Journal of Educational Measurement, 8:15-19, 1971.

Lumsden, E. A. "A Simple and Reliable Method for Assessing Partial Knowledge with Objective Tests," Journal of Instructional Psychology, 4(1):7-21.

Macintosh, H. G. (Ed.). Techniques and Problems of Assessment. London: Edward Arnold, 1974.

McGuire, C. H. "An Evaluation Model for Professional Education-Medical Education," Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1968.

Millman, J. and Pauk, W. How to Take Tests. San Francisco: McGraw-Hill, 1969.

Milton, O. and Edgerly, John W. The Testing and Grading of Students. A Change Publication, 1976.

Montgomery, R. A New Examination of Examinations. Boston: Routledge and Kegal Paul, 1978.

Nie, N., Hull, C., Jenkins, J., Steinbrenner, K., and Brent, D. SPSS: Statistical Package for the Social Sciences, 2nd Edition. San Francisco: McGraw-Hill, 1970.

Nunnally, J. C. Educational Measurement and Evaluation. New York: McGraw-Hill, 1972.

O'Neill, G. W. and Johnson, J. M. "An Analysis of Test Items as a Determinant of Student Academic Performance and Study Behavior," Journal of Personalized Instruction, 1(2):123-127, 1976.

Stanley, J. and Wang, M. D. "Weighting Test Items and Test Item Options: An Overview of the Analytical and Empirical Literature," Educational and Psychological Measurement, 30:21-35, 1970.

11. In the film on achievement, high achievers as children had parents who:

a. stressed their child's independence.
b. always underestimated their child's potential.
c. verbally "hassled" their child to higher and higher levels of performance.
d. were fairly representative of the general population.

12. High "need achievement" individuals:

a. prefer tasks of intermediate difficulty.
b. prefer tasks in which there is concrete feedback.
c. may be an important determinant for the economic growth of entire societies.
d. all of the above.

13. Cross-cultural achievement motivation training on Indian businessmen:

a. was largely unsuccessful since achievement motivation appears to be inborn.
b. was largely unsuccessful because achievement motivation seems unique to our culture.
c. was successful in terms of increased business starts, capital investment, and earnings.
d. led to the unprovoked attack on Fort Apache, Arizona.

14. One problem with need achievement training is that:

a. it does not increase motivation.
b. incerased motivation may lead to an unwanted push for organizational change.
c. it is very difficult to generate on internal standard of excellence.
d. all of the above.

15. Operant conditioning theory suggests that in order to "motivate" employees:

a. make sure that positive consequences follow the behavior you want them to perform.
b. you must first uncover the person's inner motives.
c. make sure the person only takes moderate risks.
d. provide an extensive physical education program.

16. Which of the following is an example of referent power in leadership?

a. the ability to reward and punish your subordinates
b. the charismatic personality of Ronald Reagen
c. expertise in the task at hand
d. the rights of Congress to enact laws

"MULTIPLE-CHOICE ESSAY" TEST ANALYSIS

1. Compared to exams for other courses, this exam was:

   _____ more difficult

   _____ about the same

   _____ easier

2. (a) In general, I prefer: (#1 = most preferred, #2 = next most, etc.)

   _____ multiple-choice

   _____ long essays (one or two question tests)

   _____ short essays

   _____ true-false

   _____ matching

   _____ oral exams

   _____ other _____

   (b) Why do you have this preference?

   (c) Which are the two most frequent tests you usually have?

   1. most frequent _____

   2. next frequent _____

3. How would you rate the "Multiple Choice-Essay"?

   (a) _____ Much better than my generally preferred test.
   _____ Somewhat better than my generally preferred test.
   _____ About the same as my generally preferred test.
   _____ Somewhat worse than my generally preferred test.
   _____ Much less than my generally preferred test.

   (b) _____ Much better than most multiple choice test formats.
   _____ Somewhat better than most multiple choice test formats.
   _____ About the same as most multiple choice test formats.
   _____ Somewhat worse than most multiple choice test formats.
   _____ Much less than most multiple choice test formats.

(OVER)

4. I wrote essays on _____ questions.

(a) If you wrote essays on 0 (none), did this option make the test easier for you in any way?

(b) If you wrote essays on 1 or more questions, how do you feel this option helped you?

5. Any additional comments.