DOCUMENT RESUME

ED 065 551 TM 001 696

AUTHOR Garvin, Alfred D.
TITLE Confidence Weighting Plus Coombs-Type Response Options: A Good Idea That Failed.
NOTE 7p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Confidence Testing; *Multiple Choice Tests; *Response Style (Tests); Student Testing; *Test Construction; *Test Reliability
IDENTIFIERS Confidence Weighting

ABSTRACT

Confidence weighting (CW) tends to improve the reliability of easy tests; the Coombs-type multiple-response (MR) option tends to improve the reliability of hard tests. It was hypothesized that, on a test of moderate difficulty, offering both the CW and MR response options would improve reliability more than either alone. Twenty-four subjects took a 20-item multiple-choice test under CW plus MR instructions. MR was used less than CW; 9 subjects used both options. Coefficient alphas computed on four scoring bases showed MR, alone, depressed reliability a little; CW, alone, depressed it a lot; and the two combined depressed it even more. It was concluded that these two previously successful special testing procedures cannot be combined to form an even better one. (Author)

26.9

Confidence weighting plus Coombs-type response options: A good idea that failed

Alfred D. Garvin

University of Cincinnati

Abstract

Confidence weighting (CW) tends to improve the reliability of easy tests; the Coombs-type multiple-response (MR) option tends to improve the reliability of hard tests. It was hypothesized that, on a test of moderate difficulty, offering both the CW and MR response options would improve reliability more than either alone. 24 Ss took a 20-item multiple-choice test under CW plus MR instructions. MR was used less than CW; 9 Ss used both options. Coefficient alphas computed on four scoring bases showed MR, alone, depressed reliability a little, CW, alone, depressed it a lot, and the two combined depressed it even more. It was concluded that these two previously successful special testing procedures cannot be combined to form an even better one.

Confidence weighting plus Coombs-type response options: A good idea that failed

Alfred D. Garvin

University of Cincinnati

Opponents of confidence weighting (CW), e.g., Swineford (1938, 1941) and Jacobs (1968), have argued that CW confounds academic achievement with irrelevant personality traits and operates to favor the inherently confident student over the equally knowledgable but inherently diffident one. Proponents of CW, e.g., Ebel (1965a,b) and Garvin (1969, 1972), have largely ignored this criticism, arguing instead that, to a greater or lesser degree, CW generally accomplishes what it is intended to accomplish--improve test reliability. These same two rationally orthogonal arguments have also been raised regarding the psychologically complementary Coombs-type multiple response (MR) option.

Garvin and Ralston (1970) considered the uneven success of CW and MR across different testing situations and theorized that the relative efficacy of CW and MR was a function of test difficulty: CW would "work" with easy tests by permitting extra knowledge to be displayed (and rewarded); MR would work with hard tests by permitting partial knowledge to be displayed (and rewarded). In their limited empirical test, this theory was supported: On a relatively hard course pretest taken under MR instructions by one group and CW instructions by an equivalent group, MR scores were more reliable than the corresponding conventional scores while CW scores were less reliable. They concluded that either CW or MR (but not both)

would "work" in any given testing situation, depending on test difficulty. They implicitly dismissed the possibility that neither would work.

The present study was designed to cast light on all of the foregoing propositions. In a typical group a typical test will be easy for some and hard for others. Why not provide a wide range of response options so as to elicit the extra knowledge of the better students and the partial knowledge of the poorer ones? Surely, the more opportunity to display knowledge, the more reliable the test.

Method

Subjects

The Ss were the 24 graduate education majors enrolled in the author's course in Measurement and Evaluation.

Test

The test involved was a midterm exam comprising 20 four-choice multiple choice items on basic test construction principles.

Procedure

Each S was permitted to answer each item in any one of three ways, according to his confidence in his answer. The response options available and their corresponding score contingencies were as follows:

Best answer (BA)	Simply indicate the one best answer. 1 point if right; 0 if wrong.

Confidence weighting (CW) Circle your best answer selection. 2 points if right; -2 if wrong.

Multiple response (MR) Indicate your first choice and second choice. ½ point if either is right; 0 if both wrong.

The procedures for indicating each response option and scoring such responses were carefully explained and illustrated through examples. A generous time limit was allowed for the test.

Analysis of data

All responses were scored four different ways:

BA CW and BA responses and the first choices of MR responses were all scored on a BA (1 or 0) basis.

CW CW responses were scored as 2 or -2; BA responses and the first choices of MR responses were scored as 1 or 0.

MR CW and BA responses were scored as 1 or 0; MR responses were scored as ½ if either choice was right, otherwise as 0.

CW+MR CW responses were scored as 2 or -2; BA responses were scored as 1 or 0; and MR responses were scored as ½ (if either was right) or 0.

A coefficient alpha reliability was computed for each of these four sets of scores. Rank-order correlations were computed between certain variables of interest, as explained more fully in the next section.

Results

Most of the Ss "played the game"; the distribution of Ss by response options exercised was:

BA only	2
BA+MR	2
BA+CW	11
BA+CW+MR	9
	24

Every item of the test received some special responses. Nine of the 20 items received CW responses and the other 11 received both CW and MR responses.

Most of the Ss played the game quite intelligently. In general, Ss with the highest BA scores weighted the most items and the items that they weighted were the easiest ones; Ss with the lowest BA scores gave the most second choices and they did so on the hardest items. Rank-order correlations among these variables were all in the direction implied and were significant at the .05 level.

The distribution of BA scores was symmetrical and platykurtic. The mean was 12.3; the standard deviation was 2.8. The coefficient alphas for the four methods of scoring were:

BA only	.536
CW	.408
MR	.471
CW+MR	.372

Discussion

The argument that most students are either inherently confident or inherently diffident in test-taking received little support here. When given the chance to respond in either, neither, or both of these ways in a single test, 9 out of 24 Ss did some of each. Risk-taking behavior is better explained as a rational reaction to perceived item difficulty than as an inherent personality trait.

The difficulty level of this test (on a BA-score basis) was so close to the theoretical ideal for maximum discrimination that it should be said to have been of moderate difficulty for this group. We might have expected relatively small but equal proportions of responses to have been given in the CW and the MR modes. Further, we might have expected that neither CW nor MR, alone, nor their combination would have much effect on reliability.

The actual results are interesting but dismaying. The MR option was used a little and it depressed reliability a little; the CW option was used a lot and it depressed reliability a lot. Worse still, the individual effects of these two options in depressing reliability seem to be additive when they are combined.

It remains to be seen whether a much easier test would show CW to be effective and whether a much harder test would show MR to be effective, as hypothesized. Clearly, replications of this experiment are necessary. In the meantime, we must give serious attention to the present evidence that two previously successful testing procedures do not combine to form an even better one.

References

Ebel, R. L. Confidence weighting and test reliability. Journal of Educational Measurement, June, 1965a, 2, pp. 49-57.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1965b.

Garvin, A. D. The effect of confidence weighting on variation of the error of measurement. (Doctoral dissertation, University of Maryland) Ann Arbor, Mich.: University Microfilms, 1969. NO. 69-7621.

Garvin, A. D. Confidence weighting. In S. S. Jacobs (Chm.), Alternative procedures with objective tests: An examination of three strategies. Symposium presented at the meeting of the American Educational Research Association, Chicago, April 1972.

Garvin, A. D. and Ralston, N. C. Improving the reliability of course pretests. Paper presented at the meeting of the National Council on Measurement in Education, Minneapolis, March 1970.

Jacobs, S. S. An empirical investigation of the relationship between selected aspects of personality and confidence-weighting behaviors. Unpublished doctoral dissertation, University of Maryland, College Park, Maryland, 1968.

Swineford, F. The measurement of a personality trait. Journal of Educational Psychology, 1938, 29, 289-292.

Swineford, F. Analysis of a personality trait. Journal of Educational Psychology, 1941, 29, 438-444.