DOCUMENT RESUME

ED 057 074 TM 000 911

AUTHOR	Boldt, Robert F.
TITLE	An Approximately Reproducing Scoring Scheme That Aligns Random Response and Omission.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Air Force Human Resources Lab., Lowry AFB, Colo.
REPORT NO	RB-71-43
PUB DATE	Sep 71
NOTE	10p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Confidence Testing; Guessing (Tests); *Multiple Choice Tests; *Probability Theory; *Response Style (Tests); Risk; Scores; *Scoring Formulas; Testing Problems; Test Interpretation; Transformations (Mathematics)

ABSTRACT

One formulation of confidence scoring requires the examinee to indicate as a number his personal probability of the correctness of each alternative in a multiple-choice test. For this formulation, a linear transformation of the logarithm of the correct response is maximized if the examinee reports accurately his personal probability. To equate omits scores with choice scores, the transformation can be chosen so that the score is zero if the examinee indicates complete uncertainty. If this is done, the scoring function depends on the number of alternatives. One could also align uncertainty and response omission by granting credit for omitting items, though it is felt this might be hard to explain to examinees. (Author)

RB-71-43

RESEARCH BULLETIN

AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION

Robert F. Boldt

Research sponsored by the Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base, Colorado, Contract No. F-41-609-70-C-0044. This report was issued by the Air Force as AFHRL-TM-12.

Educational Testing Service
Princeton, New Jersey
September 1971

AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION

ABSTRACT

One formulation of confidence scoring requires the examinee to indicate as a number his personal probability of the correctness of each alternative in a multiple-choice test. For this formulation, a linear transformation of the logarithm of the correct response is maximized if the examinee reports accurately his personal probability. To equate omits scores with choice scores, the transformation can be chosen so that the score is zero if the examinee indicates complete uncertainty. If this is done, the scoring function depends on the number of alternatives. One could also align uncertainty and response omission by granting credit for omitting items, though it is felt this might be hard to explain to examinees.

AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION[1]

The related problems of guessing and partial knowledge have stimulated quite a lot of consideration by test-oriented persons who are dissatisfied with the limited amount of information conveyed by the responses to multiple-choice items. One way to increase this information without increasing the amount of substantive interpretation required is to allow the examinee to indicate for each alternative the amount of uncertainty, or probability, of correctness of each alternative. In so doing, one may make the testing process more palatable in that the examinee is allowed to communicate his unsureness and hence reduce the presumed feelin[illegible] of risk and anxiety associated with marking the "best" answer--he may have very mixed feelings about the "bestness" of that answer.

It should be mentioned that while there is much interest in improving testing procedures, and confidence testing is strongly suggested by some (Shuford & Massengill, 1967), confidence testing should [illegible] [illegible]raced uncritically as an improvement. Some have reservations which stem from the fact that confidence testing requires the examinee to decide whether to take a risk and how much risk to take when making each response, as will be seen. With the usual multiple-choice testing, this decision about possible risk may be less apparent to the examinee, and, hence, the personality factors operative in the two types of testing may not be the same. Swineford (1938, 1941) has presented evidence of a relation between personality factors and risk taking in confidence testing quite apart from achievements involved.

Therefore, one should take care to ascertain that the changed operations of personality factors introduced through confidence testing do not defeat the purpose of measurement.

The present paper is not responsive to the problem of personality factors but to the treatment of omitted responses. That is, it remains usual to coordinate omissions scoring with the rest of the scoring procedures, and that is the function of this paper, at least for the confidence-testing format discussed below. This format is one in which the examinee indicates his certainty of the correctness of each alternative as a nonnegative number, and the certainties recorded must sum to specified total, such as unity in the case where they are described as being probabilities of correctness. De Finetti (1962) has raised the question as to whether when this is done, the examinees will give a response directly indicative of their personal probabilities of the correctness of the responses and has introduced some scoring functions that are maximized when the responses equal those personal probabilities (De Finetti, 1965)--the not[illegible]n bei[illegible] ra[illegible]ional man will respond honestly when such behavior optimizes his expected score. Shuford, Albert, and Massengill (1966) introduced a formalization of this no[illegible]on, called the reproducing scoring property, and have pointed out that whe[illegible]e scores only the correct response, the scoring function which is reproducing is unique and is of the form

$$S = A \log B x \quad , \tag{1}$$

where S is the item score and x is the response to the correct alternative.[2] They have taken B as 10 and A as unity when the logarithm is to the base ten and introduced the arbitrary score of minus one when x is in an interval below one hundredth (so that the scoring function will be bounded). Thus

$$S_1 = 1 + \log x \ , \qquad .01 \leq x \leq 1 \ , \quad (2)$$

in their formulation.

These choices may be overly arbitrary, however, in that no provision is made for the situation where the examinee omits the item. For example, his score on an item about whose answer he hasn't the foggiest notion should be the same whether he responds to it by telling that he knows nothing about it, or whether he omits it. He should also not expect to receive more credit for marking at random at the end of a test than the examinee who does not. To correct for omissions one might use formula (2) and assign a nonzero value to the omitted items. For example, in a four-choice test the value of

$$S_1 = 1 + \log .25$$

or about .4 is the score to be assigned to each omitted item. For 2-, 3- and 5-choice items the scores assigned to omits would be about .7, .5, and .3 respectively. If these corrections for guessing are used, they may, however, still prove unsatisfactory in that the examinee may have some difficulty understanding why points should be given for omits and might adapt some truly pathological strategy out of misunderstanding unless he thinks that omits will be physically ignored in the scoring process.

When using traditional formula scoring, one sets up the formula so that the average score under random guessing is zero, and it is suggested here that such could also be done in the confidence testing situation by appropriate choice of A and B in the scoring function. This is done by setting B equal to the number of alternatives. Then when the examinee marks that his uncertainty is 1/k where k is the number of alternatives, as he would if he is indicating no information, the score would be the same as if he omitted it in that either way the score is zero. Thus the formula

$$S_k = A\ (\log k + \log x) \tag{3}$$

takes on a zero when uncertainty is expressed and does so no matter which alternative the examinee marks. Table 1 is provided with entries aligned with a zero assignment to omits, and the value used for the constant A is

$$A = 1/(\log k)$$

which sets the upper bound of the score at unity. At the lower range of the table where x approaches zero, the value of the scoring function when x is .0 is used to keep the function bounded.

The alignment provided by adjusting the score for omits as suggested either way does not allow one to distinguish the situation where a nonchance level of uncertainty is assigned to some other alternative from one where all the responses are at the chance level. To handle this situation using only one response per item, one might score only the highest certainty rewarding the response differently when it is right than when it is wrong (Boldt, 1971). When this is done, a chance response would indicate complete certainty since the certainties must sum to one.

References

Boldt, R. F. A simple confidence testing format. Research Bulletin 71-42. Princeton, N. J.: Educational Testing Service, 1971. (Also AFHRL-TR-71-31, Technical Training Division, Air Force Human Resources Laboratory, Lowry AFB, Colorado, 1971.)

De Finetti, B. Does it make sense to speak of "good probability appraisers"? In I. J. Good (Ed.), The scientist speculates. London: Heinemann, 1962. Pp.357-364.

De Finetti, B. Methods for discriminating levels of partial knowledge concerning a test item. British Journal of Mathematical and Statistical Psychology, 1965, 18, 87-123.

Shuford, E. H., Albert, A., & Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 125-145.

Shuford, E. H., & Massengill, H. E. Individual and social justice in objective testing. Report Number SMC R-10. Lexington, Mass.: Shuford-Massengill Corporation, 1967.

Swineford, F. Measurement of a personality trait. Journal of Educational Psychology, 1938, 29, 295-300.

Swineford, F. Analysis of a personality trait. Journal of Educational Psychology, 1941, 32, 438-444.

Footnotes

[1]This research was supported by the Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base, Colorado. This report was issued by them as AFHRL-TM-12.

[2]The development of formula (1) can be carried out as follows. Let $S_h(r_h)$ be the score assigned if alternative h is correct and the examinee has indicated an amount of certainty equal to r_h. Then if p_h is his subjective probability that alternative h is correct, his expected score over all alternatives is

$$E = \sum_h p_h S_h(r_h) ,$$

and it is desired to have E at a maximum when $r_h = p_h$ subject to the constraint that

$$\sum_h r_h = 1 .$$

Thus the objective function

$$E = \sum_h p_h S_h(r_h) + A(1 - \sum_h r_h) ,$$

where A is the Lagrange multiplier imposing the condition that the r's sum to one, is maximized when

$$p_h \left. \frac{dS_h(r_h)}{dr_h} \right|_{r_h = p_h} = A,$$

or

$$\frac{dS_h(p_h)}{dp_h} = \frac{A}{p_h} .$$

Therefore, E is at a maximum when the scoring function, S, is

$$S = A \log Bx$$

where x is the indicated certainty for the correct answer, and B is a constant of integration. The proof is ancillary to the text of the paper but is included as it is quite a bit simpler than that given by Shuford et al.(1966).

Table 1

Score Certainty (%) of Correctness

of Alternative Keyed Correct

	Number of Alternatives			
Confidence	2	3	4	5
0	-56	-32	-23	-19
5	-33	-17	-12	- 9
10	-23	-11	- 7	- 4
15	-17	- 7	- 4	- 2
20	-13	- 5	- 2	0
25	-10	- 3	0	1
30	- 7	- 1	1	3
1/3	- 6	0	2	3
35	- 5	0	2	3
40	- 3	2	3	4
45	- 2	3	4	5
50	0	4	5	6
55	1	5	6	6
60	3	5	6	7
65	4	6	7	7
2/3	4	6	7	8
70	5	7	7	8
75	6	7	8	8
80	7	8	8	9
85	8	9	9	9
90	8	9	9	9
95	9	10	10	10
100	10	10	10	10