DOCUMENT RESUME

ED 107 707 TM 004 534

AUTHOR Boldt, Robert F.
TITLE An Approximately Reproducing Scoring Scheme that Aligns Random Response and Omission. Memorandum Report for Period July 1970-July 1971.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Air Force Human Resources Lab., Lowry AFB, Colo. Technical Training Div.
REPORT NO AFHRL-TR-74-99
PUB DATE Nov 74
NOTE 9p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Confidence Testing; *Guessing (Tests); *Multiple Choice Tests; *Probability; *Response Style (Tests); Scoring; *Scoring Formulas; Statistical Analysis; Testing
IDENTIFIERS Air Force

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 accurately reports 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. (Author)

AIR FORCE HUMAN RESOURCES LABORATORY

AFHRL-TR-74-99

AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION

By

Robert F. Boldt
Educational Testing Service
Princeton, New Jersey 08540

TECHNICAL TRAINING DIVISION
Lowry Air Force Base, Colorado 80230

November 1974
Memorandum Report for Period July 1970 – July 1971

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This memorandum was submitted by Educational Testing Service, Princeton, New Jersey 08540, under contract F41609-70-C-0044, project 1121, with Technical Training Division, Air Force Human Resources Laboratory (AFSC), Lowry Air Force Base, Colorado 80230. Dr. Marty R. Rockway was the project scientist and Captain Wayne S. Sellman was the task scientist.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved.

MARTY R. ROCKWAY, Technical Director
Technical Training Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFHRL-TR-74-99	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION	5. TYPE OF REPORT & PERIOD COVERED: Memorandum July 1970 – July 1971
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Robert F. Boldt	8. CONTRACT OR GRANT NUMBER(s): F41609-70-C-0044
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Educational Testing Service, Princeton, New Jersey 08540	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62703F 11210303
11. CONTROLLING OFFICE NAME AND ADDRESS: Hq Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235	12. REPORT DATE: November 1974
	13. NUMBER OF PAGES: 8
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base, Colorado 80230	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Formerly published as AFHRL (TT) TRM-12.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

psychological testing
psychometrics
confidence testing
subjective probability
technical training

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

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 accurately reports 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.

SUMMARY

Problem

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 it has been shown that a linear transformation of the logarithm of the correct alternative is a scoring function which maximizes the expected score of the examinee if he accurately reports his personal probabilities. The present paper calculates the expected score corresponding to a chance level of personal probability, thus allowing the equating of lack of response with complete uncertainty.

Approach

The solution required can be reached merely by imposing an appropriate boundary condition on the solution to a differential equation. The condition is that when the examinee indicates a certainty equal to the reciprocal of the number of alternatives, the implied score should be zero or the implied score should be awarded when a response is omitted.

Result

If one grants score points for omitted items, one may equate omission scores to chance score. However, one must explain so the examinees understand that credit may be given for omitted items. Alternatively, one may modify the scoring formula by subtracting a constant to produce a zero when a chance level of personal probability is indicated. In this case, the scoring formula is a function of the number of alternatives. However, it is possible to ascribe uncertainity to the answer when that answer is not the preferred answer; i.e., the method does not rigorously imply uncertainty.

Conclusion

No single scoring system will handle all numbers of alternatives if omission and the complete uncertainty are to be aligned. If certainty at the level of the reciprocal of the number of alternatives is rigourously to indicate complete uncertainty, the response format discussed here is not suitable.

PREFACE

The research reported in this memorandum is the result of work performed by the Educational Testing Service, Princeton, N.J. under the provisions of Contract F 41609-70-C-0044. Project Monitor was Capt Wayne S. Sellman. The research was conducted under Project 1121, Technical Training Development; Task 112103, Evaluating Individual Proficiency and Technical Training Programs. Dr. Marty R. Rockway was the Project Scientist and Capt Wayne S. Sellman was the Task Scientist.

AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALIGNS RANDOM RESPONSE AND OMISSION

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 feelings 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 and Massengill, 1967), confidence testing should not be embraced 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 and 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 scorings 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 non-negative 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 (1965)—the notion being that a rational man will respond honestly when such behavior optimizes his expected score. Shuford, Albert, and Massengill (1966) introduced a formalization of this notion, called the reproducing scoring property, and have pointed out that when one scores only the correct response, the scoring function which is reproducing is unique and is of the form

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

where S is the item score and x is the response to the correct alternative.[1] They have taken B as ten and A

---

[1] 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 an 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).

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

$$S_1 = 1 + \log x \qquad .01 \leqslant x \geqslant 1 \tag{2}$$

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 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 non-zero value to the omitted items. For example, in a four choice test the value of

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

or about .04 is the score to be assigned to each omitted item. For two, three, and five 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 l/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 non chance 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.

Table 1. **Score Certainty (%) of Correctness of Alternative Keyed Correct**

Confidence	Number of Alternatives 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

REFERENCES

Boldt, R.F. *A simple confidence testing format* AFHRL-TR-71-31, AD-737 113. Lowry AFB, CO Technical Training Division, Air Force Human Resources Laboratory, July 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, 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.