PSYCHOMETRIKA—VOL. 48, NO. 3.
SEPTEMBER, 1983
NOTES AND COMMENTS

MAXIMUM LIKELIHOOD ESTIMATION OF ITEM RESPONSE PARAMETERS WHEN SOME RESPONSES ARE OMITTED

FREDERIC M. LORD

EDUCATIONAL TESTING SERVICE

A theoretical model is given for dealing with omitted responses. Two special cases are investigated.

Key words: guessing, item response theory, latent trait theory, formula scoring, correction for guessing.

The usual item response theory (IRT) for dichotomously scored multiple-choice items cannot deal appropriately with omitted responses. It is clearly incorrect to score an 'omit' as if it were a wrong response: If we did so, the examinee who omits half the items on a true-false test and answers the remaining items correctly would be assigned a much lower estimated ability than if he has responded to all items. Assuming that the examinee has time to answer all items, as is assumed throughout this paper, it is also incorrect simply to act as if the omitted items had not been administered to the examinee at all [see Lord, 1980, Section 15.8].

Lord [1974] outlined the method used by the computer program LOGIST [Wingersky, in press; Wingersky, Barton, & Lord, Note 1] for dealing with omitted responses. This method gives asymptotically the same ability estimates that would be obtained if random responses had been recorded on the answer sheet for the omitted multiple-choice items. When the examinee responds at random to a certain item, however, he has a probability of $1/A$ of obtaining a correct answer, where A is the number of choices per item, whereas the IRT model requires that this probability be a certain increasing function of the examinee's ability. In this case, the usual IRT model cannot be correct.

Bock [1972], Samejima [1969, Notes 2, 3], and others have developed models for polychotomous items. Bock reports results when omits are simply treated as an additional response category. The present writer is uncertain of this approach, however, since it treats probability of omitting as dependent only on the examinee's ability, whereas it actually depends also on a dimension of temperament. It seems likely that unidimensional local independence may not hold. In practice, the empirical regression of number of omitted items on test score is found to take on many shapes that seem to be incompatible with the IRT assumptions made.

Section 1 considers a possible rigorous mathematical model for omitting behavior. Sections 2 and 3 show that this model yields very reasonable results in two special cases that are sufficiently simple to be already familiar.

This work was supported in part by contract N00014-80-C-0402, project designation NR 150-453 between the Office of Naval Research and Educational Testing Service. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Reprint requests should be addressed to Frederic M. Lord, Educational Testing Service, Princeton, New Jersey 08541.

0033-3123/83/0900-4118$00.50/0

1. A Model for Response or Omission

If test score is the number of right answers, an examinee who omits responses on a multiple-choice test necessarily lowers his expected test score. Mathematical modeling of such (usually irrational) behavior will not be attempted here.

We will deal instead with the case where the 'formula score' $x - W/(A - 1)$ is to be used, x being the number of right answers, W the number of wrong answers, and A the number of choices per item. In this case, the examinee who wishes to maximize his expected score will not omit an item if he feels that some of the choices are more likely to be correct than others. If he can do no better than a random guess on an item, his chance of success equals $1/A$ and his expected test score will be the same whether he omits the item or guesses at random; he will omit or guess at random as he pleases.

Let $R_i(\theta)$ denote the probability that a randomly chosen examinee at ability level θ feels no preference for any of the available responses to item i. Since he feels no preference, if he responds to item i his response is in effect at random. For brevity, we will say that $R_i(\theta)$ is the probability that the examinee is (totally) undecided on item i; $R_i(\theta)$ may be thought of as the proportion of examinees at ability level θ who will either omit item i or respond at random.

If the examinee feels any preference among the choices, he presumably will record a preference on the answer sheet. Should he fail to do this, the assumptions of the present model will be violated. Bliss [1980], Cross and Frary [1977], and other earlier writers have carried out empirical studies showing that examinees do better than chance when required to answer omitted items. Both cited studies use the same guessing directions, taken from Davis [1967]. Although these directions are typical of current practice, it is obvious that much more effective directions could be written to help (persuade) the examinee to use a guessing strategy that maximizes his score. The cited studies actually prove the need for more emphatic, less equivocal directions.

If an examinee responds to item i, denote a correct response by $u_i = 1$, an incorrect response by $u_i = 0$. If he omits item i, it will not matter what finite value is assigned to the symbol u_i in the algebra that follows. We may assign an arbitrary value to this as yet undefined symbol.

Let $P_i(\theta)$ denote the conditional probability that (a randomly chosen) examinee at ability level θ will respond correctly to item i, given that he is not (totally) undecided:

$$P_i(\theta) \equiv \text{Prob}\,(u_i = 1 \mid \text{not undecided}).$$

If the examinee omits item i, we will denote this by $w_i = 1$; if he responds to item i, $w_i = 0$. Let ω denote the probability that the examinee will omit an item on which he is (totally) undecided:

$$\omega \equiv \text{Prob}\,(w_i = 1 \mid \text{undecided}).$$

Both θ and ω are parameters that vary across examinees. (The numerical value of ω will obviously be affected by the instructions given to the examinees.)

Let $C \equiv 1/A$ denote the probability of success by random guessing:

$$C = \text{Prob}\,(u_i = 1 \mid w_i = 0, \text{undecided}).$$

Let a bar above a symbol denote its complement, for example, $\bar{P}_i \equiv 1 - P_i$.

Consider a single examinee responding to item i. There are three exhaustive and mutually exclusive events, denoted by ($w_i = 1$, u_i = arbitrary value), ($w_i = 0$, $u_i = 1$), and ($w_i = 0$, $u_i = 0$). The unconditional probability of omitting is clearly

$$\text{Prob}\,(w_i = 1) = \omega R_i. \tag{1}$$

A correct answer occurs with probability $\bar{\omega}C$ when the examinee is undecided and with probability P_i when he is not undecided, so

$$\text{Prob}\,(u_i = 1,\, w_i = 0) = R_i\bar{\omega}C + \bar{R}_i P_i. \tag{2}$$

Similarly,

$$\text{Prob}\,(u_i = 0,\, w_i = 0) = R_i\bar{\omega}\bar{C} + \bar{R}_i \bar{P}_i. \tag{3}$$

The right sides of (1), (2), and (3) sum to 1, as they should.

For the permissible values of (w_i, u_i), the joint distribution of w_i and u_i $(i = 1, 2, \ldots, n)$ may be conveniently written

$$L = \prod_{i=1}^{n} (R_i \omega)^{w_i} (P_i \bar{R}_i + C\bar{\omega}R_i)^{u_i \bar{w}_i} (\bar{P}_i \bar{R}_i + \bar{C}\bar{\omega}R_i)^{\bar{u}_i \bar{w}_i} \tag{4}$$

The log likelihood is then

$$\log L = \sum_{i=1}^{n} [w_i(\log R_i + \log \omega) + u_i \bar{w}_i \log (P_i \bar{R}_i + C\bar{\omega}R_i) + \bar{u}_i \bar{w}_i \log (\bar{P}_i \bar{R}_i + \bar{C}\bar{\omega}R_i)] \tag{5}$$

Equations for maximum likelihood estimation can be written down from (5).

The following is suggested as a possible, plausible implementation of this model:

i The parameter ω varies across examinees but not across items.

ii $P_i \equiv P_i(\theta)$ is a three-parameter logistic or normal ogive function of θ with positive slope:

$$P_i = c_i + (1 - c_i)\Phi[a_i(\theta - b_i)],$$

where Φ is the cumulative logistic or normal distribution function and a_i, b_i, and c_i are item parameters.

iii $R_i \equiv R_i(\theta)$ has the same general mathematical form as $P_i(\theta)$ but different parameters; in particular, the slope is negative, the asymptote c_i is zero.

There are five item parameters (three for P, two for R) and two examinee parameters (ω and θ). This plausible implementation will be assumed in all that follows.

Some special cases of the model are considered in the remaining sections. All item parameters are assumed to be known. The purpose is to gain further insight into the implications of the model.

2. *Special Case: No Omitting*

When $\omega = 0$, with probability 1 we must have $w_i = 0$, and (4) becomes

$$L = \prod_{i=1}^{n} p_i^{u_i}(1 - p_i)^{1-u_i} \tag{6}$$

where

$$p_i \equiv P_i(\theta) \equiv P_i \bar{R}_i + CR_i. \tag{7}$$

We have here the usual item response theory model for dichotomous items, omits being barred, except that now the item response function has the special form given by (7).

This item response function $p_i(\theta)$ need not be a monotonic increasing function of θ. The lower asymptote of $p_i(\theta)$ is $p_i(-\infty) = C$; if c_i, the lower asymptote of $P_i(\theta)$, is less than C, $p_i(\theta)$ may decrease at first as θ increases, before finally increasing to the upper asymptote at 1. This is a desirable feature: Examinees with sufficiently low θ can only guess randomly, examinees with higher θ may be misinformed or attracted by crafty distractors and may do less well than a random guess. Model (6)–(7) was suggested by Samejima [Note 2].

3. *Special Case: Equivalent Items, Knowledge or Random Guessing*

Under the knowledge-or-random-guessing assumption, the examinee either knows the answer to a particular item or guesses at random or omits it. For purposes of the present model, the assumption is represented by the case where $P_i(\theta) = 1$ for all θ. It follows that $\bar{R}_i(\theta)$ is the probability that an examinee at ability θ knows the correct answer to item i.

When the test is composed of equivalent items, the subscript i can be dropped from functions of parameters. Denote the number of omitted items by $V \equiv \sum_i w_i$, the number of wrong answers by $W = \sum_i \bar{u}_i \bar{w}_i$, and the number of right answers by $x \equiv n - V - W$.

The log likelihood (5) is now

$$\log L = V \log R + V \log \omega + x \log (\bar{R} + C\bar{\omega}R) + W \log \bar{C} + \log \bar{\omega} + \log R). \tag{8}$$

If we differentiate this with respect to ω and set the result equal to zero, we obtain the likelihood equation

$$\frac{V}{\hat{\omega}} - \frac{Cx\hat{R}}{\hat{\bar{R}} + C\hat{\bar{\omega}}\hat{R}} - \frac{W}{\hat{\bar{\omega}}} = 0 \tag{9}$$

where $\hat{\omega}$ and $\hat{R}$ denote maximum likelihood estimators.

The likelihood equation for θ is seen to be

$$\frac{\partial \log L}{\partial \theta} \equiv \frac{\partial \log L}{\partial R} \frac{\partial R}{\partial \theta} = 0$$

or simply $\partial \log L/\partial R = 0$, since by assumption 3, $\partial R/\partial \theta \neq 0$. It is convenient for some purposes to think of R itself as the ability parameter, since $R \equiv R(\theta)$ is a one-to-one monotonic transformation of the parameter θ.

The remaining likelihood equation is thus seen to be

$$\frac{V}{\hat{R}} + \frac{x(-1 + C\hat{\bar{\omega}})}{\hat{\bar{R}} + C\hat{\bar{\omega}}\hat{R}} + \frac{W}{\hat{R}} = 0. \tag{10}$$

The maximum likelihood estimators of ω and of θ (or R) are the roots of (9) and (10).

Rewrite (9) and (10):

$$\frac{V}{\hat{\omega}} - \frac{W}{\hat{\bar{\omega}}} = \frac{Cx}{C\hat{\bar{\omega}} + \dfrac{\hat{\bar{R}}}{\hat{R}}}, \tag{11}$$

$$V + W = \frac{x(1 - C\hat{\bar{\omega}})}{C\hat{\bar{\omega}} + \dfrac{\hat{\bar{R}}}{\hat{R}}}. \tag{12}$$

Eliminating $\hat{R}$ from these two equations, we have

$$V + W = \left(\frac{V}{\hat{\omega}} - \frac{W}{\hat{\bar{\omega}}}\right)\left(\frac{1}{C} - \hat{\bar{\omega}}\right).$$

Clearing fractions, we find the maximum likelihood estimator

$$\hat{\omega} = \frac{V}{\dfrac{V + W}{1 - C}}. \tag{13}$$

Solve (12) to obtain

$$\frac{\hat{\hat{R}}}{\hat{R}} = \frac{x(1 - C\hat{\hat{\omega}})}{V + W} - C\hat{\hat{\omega}},$$

$$\frac{1}{\hat{R}} = \frac{n}{n - x}(1 - C\hat{\hat{\omega}}).$$

Using (13), we obtain the maximum likelihood estimator of the examinee's ability:

$$\hat{R} = \frac{1}{n}\left(V + \frac{W}{1 - C}\right). \tag{14}$$

From (13) and (14),

$$\hat{\omega} = \frac{V}{n\hat{R}}, \tag{15}$$

a very reasonable result. It says that the estimated probability of omitting equals the actual number of omits divided by the estimated number of items on which the examinee is totally undecided.

Similarly, from (14),

$$n\hat{\hat{R}} = x - \frac{W}{A - 1}, \tag{16}$$

where $A \equiv 1/C$. This shows that the estimated number of items known by the examinee is given by the usual 'correction for guessing' formula.

REFERENCE NOTES

1. Wingersky, M. S., Barton, M. A., & Lord, F. M. *LOGIST user's guide.* Princeton, N.J.: Educational Testing Service, 1982.
2. Samejima, F. *A new family of models for the multiple-choice item* (Research Report 79-4). Knoxville, Tenn.: Department of Psychology, University of Tennessee, 1979.
3. Samejima, F. *Final report: Efficient methods of estimating the operating characteristics of item response categories and challenge to a new model for the multiple-choice item.* (Final Report). Knoxville, Tenn.: Department of Psychology, University of Tennessee, 1981.

REFERENCES

Bliss, L. B. A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students. *Journal of Educational Measurement*, 1980, *17*, 147–153.

Bock, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 1972, *37*, 29–51.

Cross, L. H. & Frary, R. B. An empirical test of Lord's theoretical results regarding formula-scoring of multiple-choice tests. *Journal of Educational Measurement*, 1977, *14*, 313–321.

Davis, F. B. A note on the correction for chance success. *Journal of Experimental Education*, 1967, *36*, 42–47.

Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 1974, *39*, 247–264.

Lord, F. M. *Applications of item response theory to practical testing problems.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17*, 1969.

Wingersky, M. S. LOGIST: a program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *ERIBC Monograph on Applications of Item Response Theory*. Vancouver, British Columbia: Educational Research Institute of British Columbia, in press.

Manuscript received 5/28/82
First revision received 9/21/82
Final version received 4/6/83