The Item Log-Likelihood Surface for Two- and Three-Parameter Item Characteristic Curve Models

Frank B. Baker
University of Wisconsin

This article investigated the form of item log-likelihood surface under two- and three-parameter logistic models. Graphs of the log-likelihood surfaces for items under two-parameter and three-parameter (with a fixed value of c) models were very similar, but were characterized by the presence of a ridge. These graphs suggest that the task of finding the maximum of the surface should be roughly equivalent under these two models when c is fixed in the three-parameter model. For two items, the item log-likelihood surface was plotted for several values of c to obtain the contour line of the maxima. For an item whose value of Lord's $b - 2/a$ index was less than the criterion value, the contour line was relatively flat. The item having an index value above the criterion value had a contour line with a very sharp peak. Thus, under a three-parameter model, finding the maximum of the item log-likelihood is more difficult when the criterion for Lord's index is not met. These results confirm that the LOGIST program procedures used to locate the maximum of the likelihood function are consistent with the form of the item log-likelihood surface. *Index terms: estimation, item parameter; likelihood surfaces; LOGIST procedures; log-likelihood; maximum likelihood estimation.*

An important aspect of item response theory (IRT) is the use of maximum likelihood to estimate test item parameters. Although this procedure generally works well, estimating item parameters under the three-parameter item characteristic curve (ICC) model has been somewhat troublesome. The three-parameter model requires large sample sizes compared to the two-parameter model (Hulin, Lissak, & Drasgow, 1982), and the standard errors of the parameter estimates can be very large (Thissen & Wainer, 1982). In addition, estimating the so-called guessing parameter is difficult under certain circumstances. For example, Kolen (1981) reported that in three tests, the LOGIST computer program (Wingersky, Barton, & Lord, 1982) failed to yield maximum likelihood estimates (MLEs) of the guessing parameter for 92%, 53%, and 36% of the items.

Despite the well-recognized estimation problems of the three-parameter model, little has been published identifying the basis of these difficulties (Baker, 1987). Under IRT, there are h parameters for each of n items in a test and an ability (θ) parameter for each of N examinees in the group tested. The usual MLE procedures involve finding estimates of the $hn + N$ parameters that maximize the likelihood function. Because of the dimensionality of the equations, a direct solution is beyond the capabilities of most digital computers. As a result, the joint maximum likelihood estimation (JMLE) procedure first proposed by Birnbaum (1968) is widely used.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 12, No. 4, December 1988, pp. 387–395

0146-6216/88/040387-09$1.70

The JMLE approach assumes that the items and examinees are pair-wise independent over all items and examinees. Thus, the $hn + N$ simultaneous equations are reduced to n sets of h equations for the test items, and one equation for each of the examinees. Two constraints are imposed to fix the metric. These equations are embedded within a two-stage iterative process that alternatively solves for item and then examinee parameters until the maximum of the likelihood function is reached. Achieving the overall maximum is accomplished by finding the joint maximum of the likelihoods of the n items and the N examinees. Thus, when estimating item parameters, the basic task is finding the maximum of the likelihood function with respect to the h parameters of a single item under the assumption that the θ parameters of the N examinees are known.

This article focuses on the form of the likelihood surface for a single item rather than the overall likelihood function. Clearly, this is only part of the overall maximization process implemented by the JMLE procedure. However, it is a crucial part because problems in this procedure, such as those associated with the three-parameter model, have an impact on the overall solution.

The Likelihood Function

Assume that J groups of f_j examinees have θ scores θ_j ($j = 1, 2, \ldots, J$). Of the f_j examinees in the group having θ level θ_j, r_{ij} gave a correct response to item i ($i = 1, 2, 3 \ldots, n$) and $f_j - r_{ij}$ gave an incorrect response. The observed proportion of correct responses to item i at θ_j is given by

$$p_i(\theta_j) = p_{ij} = \frac{r_{ij}}{f_j} \tag{1}$$

and

$$q_{ij} = 1 - p_{ij} = \frac{f_j - r_{ij}}{f_j} \quad . \tag{2}$$

Assuming the p_{ij} are binomially distributed with $E(p_{ij}) = P_{ij}$ and variance $P_{ij}Q_{ij}/f_j$, the probability of the r_{ij} over all items and groups is given by the likelihood function

$$L = \prod_{i=1}^{n} \prod_{j=1}^{J} \frac{r_{ij}!}{r_{ij}!(f_j - r_{ij})!} P_i(\theta_j)^{r_{ij}} Q_i(\theta_j)^{f_j - r_{ij}} \quad . \tag{3}$$

The $P_i(\theta_j)$ are functions of the item parameters under the specified ICC model and the examinee θ levels. For a given item, under a three-parameter model,

$$P_i(\theta_j) = P_j = c + (1 - c)\{1 + \exp[-1.702\, a(\theta_j - b)]\}^{-1} \quad , \tag{4}$$

where a is the discrimination parameter,
b is the difficulty parameter,
c is the so-called guessing parameter, and
1.702 is used to approximate the normal ogive metric.

This article focuses on the likelihood function for a single item, which is

$$L = \prod_{j=1}^{J} \frac{r_j!}{r_j!(f_j - r_j)!} P_j^{r_j} Q^{f_j - r_{ij}} \quad . \tag{5}$$

The logarithm of the likelihood function for an item is

$$\mathrm{Log}(L) = \sum_{j=1}^{J} \mathrm{Log}\left[\frac{r_j!}{r_j!(r_j - f_j)!}\right] + \sum_{j=1}^{J} r_j \mathrm{Log}(P_j) + \sum_{j=1}^{J} (r_j - f_j)\mathrm{Log}(Q_j) \quad . \tag{6}$$

Lord (1980) provided the likelihood equations solved for the item parameter estimates. Because of the estimation difficulties inherent in the three-parameter model, the LOGIST computer program imposes a complex set of restrictions both within and across stages of the JMLE paradigm (Wingersky, 1983). Without these restrictions, the iterative process for a given item often diverges rather than converges and infinite estimators result.

The need to impose restrictions on the item parameter estimation process suggests that the form of the log-likelihood surface is not conducive to easily finding its maximum. For example, Lord (1980, p. 181) indicated that parameter estimates have a large sampling variance when the likelihood function has a relatively flat maximum. Thissen and Wainer (1982) showed that the asymptotic variances are very large under a three-parameter model. Such results can occur when the likelihood function has a relatively flat overall surface or when it has a ridge with a relatively flat top.

Under a two-parameter model, the MLE procedure works better and generally converges readily if it has good initial estimates (Baker, Harwell, & Serlin, 1982). This suggests that the form of the likelihood surface for the two ICC models differs and that the form of the three-parameter model surface may be suspect. Hulin et al. (1982) said regarding the three-parameter model that "the likelihood surface is not well suited to quadratic methods of function maximization. It seems reasonable to speculate that the difficulties are caused more by the likelihood surface than by the method of function maximization" (p. 259).

Only Warsavage (1983) has dealt directly with the form of the likelihood surface. She plotted the item-likelihood surface for two items taken from the Reading Vocabulary subtest of the Comprehensive Tests of Basic Skills (form U, Level J) administered to 2,804 examinees. The graphs of the likelihood functions for these two items (a and b were allowed to vary while the value of c was fixed) showed a sharply peaked surface. Her results did not reveal anything unusual, but were limited by the small number of items and item parameter values investigated. She also did not graph the log-likelihood function. Although the maxima of the likelihood and log-likelihood surfaces occur at the same values of the item parameters, the forms of the surfaces will be different. Because the MLE procedure employs the latter, the present paper examines only the form of the item log-likelihood surface.

Method

Graphs of the item log-likelihood surface were obtained for eight items (four each under a two- and three-parameter model). The items under the two models were paired on the basis of the values of the discrimination parameter a and the difficulty parameter b. Under the three-parameter model the LOGIST program employs Lord's $b - 2/a$ index to determine if the c parameter should be estimated or set to a fixed value. This index corresponds roughly to the point on the θ scale where the ICC is within .03 of the value of c, so that the ICC is becoming asymptotic to c (Wingersky et al., 1982). For samples of 2,000 to 3,000, Wingersky et al. suggested a criterion value of -3.5. When the value of $b - 2/a$ is less than -3.5, the c parameter is not estimated because not enough data points exist to provide information about the value of c. In this case, the LOGIST program substitutes the average value of c yielded by test items in which c could be estimated and uses this fixed value, $\tilde{c}$, in the Newton-Raphson procedure.

This article attempts to determine whether the form of the log-likelihood surface differs among items that meet or do not meet this criterion. In addition, it is important to determine whether these surfaces differ from those obtained under a two-parameter model. Because the graph of the log-likelihood under a three-parameter model is a four-dimensional figure, the dimensionality needs to be reduced. A fixed value of $c = .30$ was used to achieve a three-dimensional graphing space and to obtain an ICC that clearly differs from that of a two-parameter model with the same values of a and b. In computing the log-likelihood surface, only a and b were allowed to vary. The goal was to determine if the log-likelihood surfaces differed under the two models.

Within the set of eight items, the item parameter values were selected to provide three levels of discrimination and difficulty as well as combinations of values that did and did not meet a criterion of -3.5 for Lord's index. Item 1A ($a = .57$, $b = -2.5$) and item 1C ($a = .57$, $b = -2.5$, $c = .3$) were easy items with moderate discrimination. Item 1C did not meet Lord's criterion for estimating the c parameter as $b - 2/a = -6$. Item 2A ($a = .36$, $b = -.5$) and item 2C ($a = .36$, $b = -.5$, $c = .3$) were moderately easy items with low discrimination. Item 2C did not meet Lord's criterion as $b - 2/a = -6$. Item 3A ($a = .8$, $b = -1$) and item 3C ($a = .8$, $b = -1$, $c = .3$) were easy items with moderately high discrimination. The value of $b - 2/a = -3.5$ for item 3C is at Lord's criterion. Item 4A ($a = 1$, $b = 2$) and item 4C ($a = 1$, $b = 2$, $c = .3$) were difficult items with high discrimination. The value of $b - 2/a = 0$ for item 4C met Lord's criterion.

The first step in creating a log-likelihood surface is determining the values of r_j and $f_j - r_j$ for a given item as a function of the J θ levels employed and the item's parameter values. Normally, the r_j would be obtained from the f_j examinees' dichotomously scored item responses at each of the J θ levels, but they were simulated here. Because f_j and r_j appear in the exponents and in the factorial terms of Equation 6, the form of the log-likelihood surface is sensitive to the location of the ICC relative to the distribution of examinees over the θ scale. Because of this, a uniform distribution of examinees was employed. Such a distribution has been employed in many IRT investigations (e.g., Ree, 1979; Swaminathan & Gifford, 1983; Wingersky & Lord, 1984).

For a given item, the ICC specified by the parameter values was evaluated at 33 levels of θ, where $-4 \leq \theta \leq 4$, yielding P_j and $r_j = f_j P_j$. The value of f_j was set to 100 at the 33 θ levels so the overall sample size was 3,300. Because this article focuses on the portion of the log-likelihood surface containing the maximum, the log-likelihood for an item was evaluated for 21 values of a and 21 values of b centered around the item's parameter values. In the case of the three-parameter model, the value of c was not varied.

Thus, each surface consisted of 441 points. Because the value of the log-likelihood is negative, the value $200 + L$ was used to obtain surfaces having a positive maximum, which facilitated plotting. In each evaluation of the log-likelihood function the "true" values of f_j, r_j, and $f_j - r_j$ for the given items were used and remained fixed while a and b were varied. Thus, the plotted figures represent the log-likelihood surfaces that occurred when the uniformly distributed examinees responded to the item exactly in the manner specified by the item's parameters, making it in some sense an optimal surface.

Results

The item log-likelihood surfaces for the eight items are presented in Figure 1 with the two-parameter surfaces in the left column. The individual figures have been rotated 30 degrees toward the readers to better show the surface, which might make the maxima less obvious. In Figure 1 the value of c in the three-parameter model was fixed. Thus, the graphs do not represent the surface over which the maximum likelihood process must move when a, b, and c are estimated simultaneously. To represent the four-dimensional graph, the following approach was used: The 33 values of r_j for item 2 were computed from the item parameter values $a = .36$, $b = -.5$, $c = .15$, and the uniformly distributed examinees. A value of $c = .15$ was used because it is a value of $\hat{c}$ commonly yielded by LOGIST when this parameter cannot be estimated.

The resulting response vector $\mathbb{R} = (r_1, r_2, \ldots, r_{33})$ was used when computing the log-likelihood surface for $c = 0$, $c = .07$, $c = .15$, $c = .23$, and $c = .30$. For each value of c, the values of the parameters a and b were varied. Due to the visual complexity of the overlaid graphs, only three of the five surfaces were plotted on the same a, b, L coordinate system in Figure 2. However, the maxima of all five surfaces are connected by a heavy line to show the trajectory of the maximum log-likelihood as a function of the

Figure 1
Log-Likelihood Surfaces for Two-Parameter
and Three-Parameter (Fixed *c*) Models

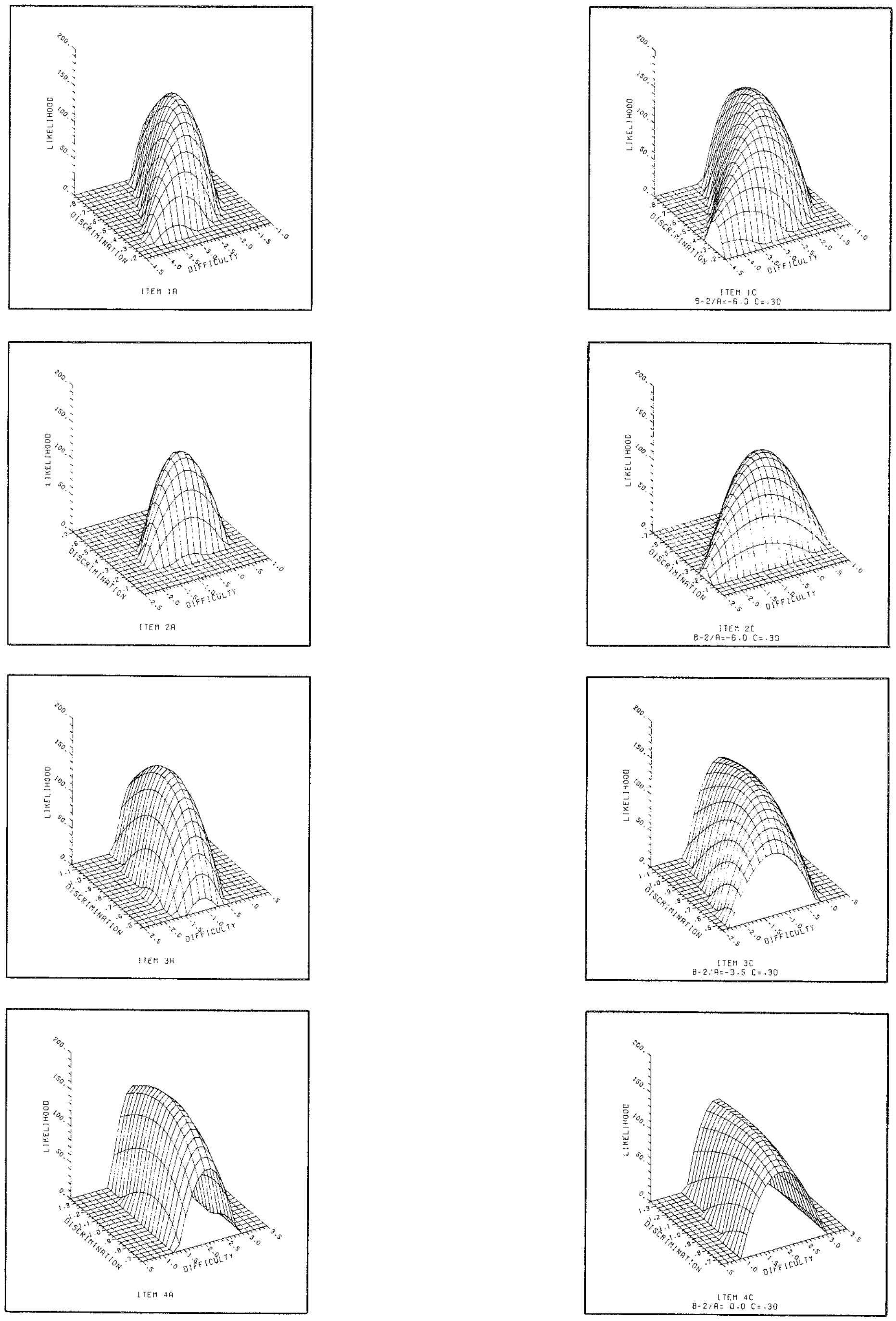
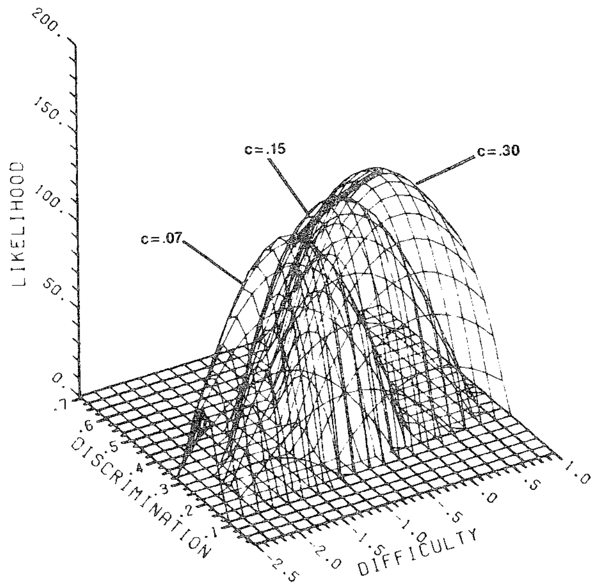


Figure 2
Log-Likelihood Surfaces for Item 2



value of c, which yields a rough idea of the effect of the c parameter. The locations of the maxima in the parameter space are given in Table 1.

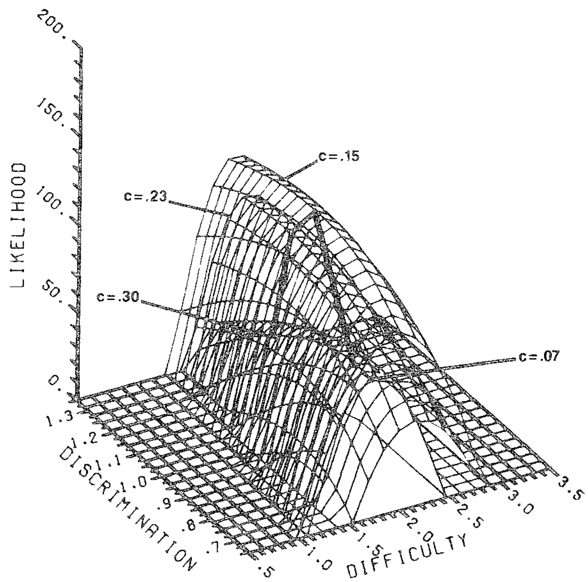
A similar procedure was employed for item 4. The response vector ℝ for item 4 was computed from the item parameters $a = 1$, $b = 2$, $c = .15$, and the uniformly distributed examinees. This response vector was used when the log-likelihood surface was computed for $c = .07$, $c = .15$, $c = .23$, and $c = .30$. To retain the coordinate system used above, the likelihood surface for $c = 0$ was not used because it had a rather large negative maximum. The surfaces for the first three values of c are shown in Figure 3 and the maxima corresponding to all four values of c are connected by a heavy line. The locations of the maxima in the parameter space also are reported in Table 1.

Figure 1 shows that the forms of the log-likelihood surfaces were similar under the two-parameter and three-parameter models when the value of c is fixed under the three-parameter model. Fixing c appears to spread out the log-likelihood surface and create a greater bend in the major axis of the surface. All of the eight surfaces had a ridge-like appearance, but the sharpness and orientation of the ridge seems to be a function of the value of Lord's $b - 2/a$ index.

Table 1
Maximum of Log-likelihood as a Function of the Item Parameters for Items 2 and 4

Item 2				Item 4			
a	b	c	Maximum	a	b	c	Maximum
.25	−1.28	.00	120.36				
.32	−.93	.07	122.56	.72	1.70	.07	79.24
.36	−.50	.15	123.07	1.00	2.00	.15	126.75
.39	−.05	.23	121.29	1.17	2.15	.23	92.70
.46	.30	.30	114.47	1.28	2.30	.30	18.25

Figure 3
Log-Likelihood Surfaces for Item 4



In items having large negative values of the index (1A, 1C, 2A, and 2C) the top of the ridge was parabolic in form with a rather sharp peak, but the major axis of the log-likelihood surface had a 90-degree bend under both ICC models. At low values of discrimination, finding the maximum of the ridge depends on a, although at high levels of discrimination it depends on b.

The contour lines of the log-likelihood surface parallel to the difficulty axis or to the discrimination axis were parabolic with well-defined peaks. This suggests that finding the maximum of the log-likelihood with respect to the a and b parameters should be easy when the criterion for the index is not met.

As the value of Lord's index increased across items, a general straightening of the major axis occurred until it was virtually parallel to the discrimination axis. Items for which the values of Lord's index were equal to or more positive than the criterion value (3A, 3C, 4A, and 4C) exhibited only a small amount of bend in the major axis of the log-likelihood surface. However, a distinct difference existed in the forms of the surfaces under the two ICC models in that the contour lines parallel to the discrimination axis curved more under the two-parameter model. This suggests that finding the maximum with respect to the item discrimination parameter for the three-parameter model will be difficult even when the criterion for Lord's index is met.

The eight graphs presented in Figure 1 also show that the form of the log-likelihood surfaces is well behaved with respect to the item difficulty parameter. In all cases, the contour lines parallel to the difficulty axis were parabolic in form and had well-defined peaks, especially near the maximum of the surface. Thus, finding the maximum with respect to the difficulty parameter should be easier than with respect to the discrimination parameter under both models whether or not Lord's index is met.

Figures 2 and 3 show the log-likelihood surfaces as a function of the a, b, and c parameters. The amount of curvature in the path taken by the maxima of the log-likelihoods with respect to c appears to be a function of the value of Lord's $b - 2/a$ index. Item 2 had an index value of -6, and it is clear that the path of the maximum values of the log-likelihood surfaces for this item was quite flat with respect to the values of c. In this item, the maxima with respect to a and b would be readily achieved, but the maximum with respect to c would be elusive. Item 4 had an index value of 0 and the contour line of the maxima of the log-likelihoods, as a function of c, was parabolic with a well-defined peak. In this item, the maximum with respect to c and b should be relatively easy to locate, although it may remain elusive with respect to a. The distinct difference in the curvature of the paths of the maxima of the log-likelihood for these two items suggests that the value of Lord's index does identify items in which the estimation of the c parameter will be difficult.

Discussion

Estimating item parameters under maximum likelihood procedures involves two tasks. First, it is essential to get within the convergence neighborhood of the maximum of the likelihood function (Kale, 1962), which requires appropriate initial estimators. Second, it is necessary to converge upon the maximum of the log-likelihood function. If the surface is relatively flat in the convergence neighborhood, changes in the value of the log-likelihood are small compared to changes in the item parameter estimates. The convergence criterion then becomes a critical issue. A tight criterion coupled with large increments in the item parameters will lead to "wandering" on the log-likelihood surface and convergence will be very slow. A loose criterion will lead to convergence at an inappropriate value of the parameters. When the log-likelihood surface has a well-defined peak, small changes in the parameter values lead to large changes in the log-likelihood function and the maximum can be found easily.

The problems in finding the maximum of both the overall and item log-likelihoods were recognized early by the authors of the LOGIST computer program. Constraints were imposed on the size of the increments to the item parameter estimates within each iteration of the MLE procedures, and upper and lower bounds were set on the parameter estimates themselves. Within stages, these constraints prevent large movements of the parameter estimates.

To facilitate finding the overall maximum, a set of steps was established across the two stages of the JMLE procedure so that some parameters were fixed while others were estimated (Wingersky, 1983). In Step 1, θ and b were estimated stage-wise while the remaining parameters were fixed. In Step 2, θ was fixed while a, b, c, and $\tilde{c}$ were estimated. Step 3 repeated Step 1. In Step 4, θ and $\tilde{c}$ were fixed while a, b, and c were estimated.

Finding the maximum with respect to b in the first step will locate the ridge in the item log-likelihood surface. Once on the ridge, fixing θ and estimating all the item parameters moves the solution along the surface. Then, estimating b in Step 3 repositions the solution on the ridge. Step 4 once more moves the solution over the surface. These four steps are repeated until the overall log-likelihood is maximized. Although Lord's writings do not mention plotting the likelihood surfaces, clearly he understood their characteristics. The present results tend to confirm that the procedures implemented in the LOGIST program facilitate finding the maxima of the item log-likelihoods and hence that of the overall log-likelihood.

References

Baker, F. B. (1987). Methodology review: Item parameter estimation under the one-, two-, and three-parameter logistic models. *Applied Psychological Measurement, 11*, 111–141.

Baker, F. B., Harwell, M. R., & Serlin, R. C. (1982). *The problem of initial values in item parameter estimation under item response theory*. Unpublished manuscript, University of Wisconsin, Madison.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement, 6*, 249–260.

Kale, B. K. (1962). On the solution of likelihood equations by iteration processes. The multiparameter case. *Biometrika, 49,* 479–486.

Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement, 18,* 1–12.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.

Ree, M. J. (1979). Estimating item characteristic curves. *Applied Psychological Measurement, 3,* 371–385.

Swaminathan, H., & Gifford, J. A. (1983). Estimation of parameters in the three-parameter latent trait model. In D. Weiss (Ed.), *New horizons in testing*. New York: Academic Press.

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika, 47,* 397–412.

Warsavage, B. (1983). *An analysis of the shape of the likelihood function of the three parameter logistic item response curve*. Paper presented at the annual meeting of the American Educational Research Association. Montreal.

Wingersky, M. S. (1983). LOGIST: a program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide*. Princeton NJ: Educational Testing Service.

Wingersky, M. S., & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement, 8,* 347–364.

Author's Address

Send requests for reprints or further information to Frank B. Baker, Department of Educational Psychology, Educational Sciences Building, 1025 West Johnson St., Madison WI 53706, U.S.A.