# DOCUMENT RESUME

ED 143 671 TM 006 113

| | |
|---|---|
| AUTHOR | Levine, Michael V.; Saxe, David H. |
| TITLE | The Use of Periodic Functions to Measure the Difficulty of Aptitude Test Items. |
| INSTITUTION | Educational Testing Service, Princeton, N.J. |
| REPORT NO | ETS-RB-76-17 |
| PUB DATE | Oct 76 |
| NOTE | 56p. |
| EDRS PRICE | MF-$0.83 HC-$3.50 Plus Postage. |
| DESCRIPTORS | *Aptitude Tests; *Item Analysis; Test Reliability |
| IDENTIFIERS | *Item Characteristics Curve Theory; Periodic Functions |

ABSTRACT

A novel use of periodic functions, called the periodic procedure, was recently introduced to make practical the use of certain physical measurement ideas in psychology. This paper reports a successful attempt to apply the periodic procedure to an important psychological measurement problem, the measurement of aptitude test item difficulties. The difficulties of items on a subtest of the Scholastic Aptitude Test were measured with both the periodic procedure and with an item characteristic curve theory procedure. The principal finding was a close agreement between the two methods. This close agreement was particularly striking because different samples (from the same test administration) were used by the two methods. The statistical development of the periodic procedure is presented along with an example of its use. (Author/JKS)

RB-76-17

RESEARCH BULLETIN

# THE USE OF PERIODIC FUNCTIONS TO MEASURE THE DIFFICULTY OF APTITUDE TEST ITEMS

Michael V. Levine

and

David H. Saxe



This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the authors. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
October 1976

The Use of Periodic Functions to Measure the Difficulty of Aptitude Test Items


ABSTRACT

Two different procedures were used to measure the difficulties of some Scholastic Aptitude Test items: a new distribution-free procedure that uses periodic functions and LOGIST, a well-developed optimization procedure that fits a logistic model. Despite the fact that the two procedures used different data in very different ways, they obtained virtually the same numbers.

The Use of Periodic Functions to Measure the Difficulty of Aptitude Test Items*

## INTRODUCTION

A novel use of periodic functions, called the "periodic procedure," was recently introduced (Levine, 1975) to make practical the use of certain physical measurement ideas in psychology. This paper reports a successful attempt to apply the periodic procedure to an important psychological measurement problem, the measurement of aptitude test item difficulties.

The difficulties of items on a subtest of the College Board's Scholastic Aptitude Test were measured with both the periodic procedure and with LOGIST, a well-developed alternative procedure. Our principal finding was a close agreement between the two methods. This close agreement is particularly striking because different samples (from the same test administration) were used by the two methods.

The agreement provides support in favor of both methods. On the one hand, since the periodic procedure is distribution free, the close agreement shows that the stronger assumptions underlying LOGIST are not a source of error in connection with its application to the SAT (see also Lord, 1970). On the other hand, the close agreement indicates that our weak assumptions are strong enough to determine the SAT item difficulties,

*We gratefully acknowledge the editorial advice and criticism of Joseph Kruskal, Ingram Olkin and Thomas Stroud. The long exposition of the periodic procedure in Section II was written in response to Kruskal's detailed comments on Levine (1975). This project would not have been attempted had it not been for Frederic M. Lord's earlier findings, especially Lord (1970).

and that our novel method is reasonably safe against the unanticipated difficulties that often beset new methods.

The periodic procedure exemplifies a new approach to measurement, characterized by a use of general functional equations and group theoretical methods in place of specific distribution assumptions (Levine, 1970, 1972, 1975, 1976). A great deal of work remains to be done before the new methods can be regarded as fully comparable to the better-developed, widely used "optimizations methods," i.e., to the methods which, like LOGIST, work by optimizing an index of goodness of fit for a model making a specific distribution assumption. However, the close agreement suggests that the new methods can be as accurate as the optimization methods. Furthermore, at least in some applications, functional equations based methods may eventually be preferred to optimization methods because (1) they do not require simultaneous estimation of ability parameters, which may have large sampling errors, (2) they are not iterative or otherwise susceptible to convergence problems, and (3) they rest on general psychometric assumptions and in particular are distribution free.

## I. OVERVIEW

The goal of this research is to evaluate the periodic procedure by attempting to estimate item difficulties for a subtest of the Scholastic Aptitude Test. The periodic procedure is reviewed in Section II. Some necessary psychometric theory and results are presented in Sections III and IV. Our results are given in Section V. The computer programs are discussed in Section VI.

## II. THE PERIODIC PROCEDURE

In this section some basic definitions are stated and the periodic procedure is described in a form convenient for the application. For a more general discussion of the periodic procedure, a discussion of the relationship to length measurement and a worked example, see Levine (1975).

### II.1 Definitions

The basic operation of the periodic procedure is function composition. If f and g are functions and if for all x such that g(x) is defined f[g(x)] is also defined, then the composition of f and g is defined to be the function written fg and given by the formula

$$fg(x) = f[g(x)] \quad .$$

If f is a real function with domain of definition D, the graph of f is the subset of the plane

$$\mathcal{G}(f) = \{ \langle x,y \rangle \in \mathbb{R}^2 : x \in D \text{ and } y = f(x) \} \; .$$

For our purposes, the most important fact about graphs is that the graph of the composition of two functions can be determined from the graphs of the composed functions. Thus

$$\mathcal{G}(fg) = \{ \langle x,y \rangle : \text{for some } z, \langle x,z \rangle \in \mathcal{G}(g) \;\&\; \langle z,y \rangle \in \mathcal{G}(f) \} \quad .$$

If the compositions wu and uw are both defined and if for all values of x in the domain of definition of u and y in the domain of

$w$, we have both $wu(x) = x$ and $uw(y) = y$, then $w$ will be called the <u>inverse of</u> $u$ and written $u^{-1}$. When inverses and composites are defined as above, elementary arguments can be used to show

(i) a function has an inverse if and only if it is $1 - 1$.

(ii) the inverse of a function is unique.

(iii) if $u$ is a $1 - 1$ function with inverse $w$, then $w$ is a $1 - 1$ function with inverse $u$.

(iv) if $u$ is a $1 - 1$ function with domain $D$, the graph of the inverse of $u$ is $\{ < u(x), x > : x \text{ is in } D\}$.

The last point is especially useful. In the psychometric application, we work with piecewise-linear tabled functions. The last point implies that we obtain the graph of the inverse of a function simply by interchanging $x$ and $y$ tables.

A <u>translation</u> is a real function of the form $\ell(x) = x + k$. A function $f$ is called a <u>conjugate of a translation</u> if there is a translation and a continuous increasing $1 - 1$ function $u$ such that $f = u^{-1}\ell u$, i.e., if for some constant $k$ and continuous, strictly increasing function $u$, $f(\cdot) = u^{-1}[u(\cdot) + k]$.

By repeatedly composing a function with itself, one defines the <u>iterates</u> of the function. Thus $f^2 = ff$ is the second iterate, $f^3 = ff^2$ is the second, $f^4 = ff^3$ the fourth, etc. If $f$ is conjugate to a translation, then all of its iterates are defined, and from the conjugacy equation $f(\cdot) = u^{-1}[u(\cdot) + k]$ one obtains the equation $f^n(\cdot) = u^{-1}[u(\cdot) + nk]$ for each iterate. If the usual conventions $f^0(x) = x$, $f^1 = f$ and $f^n = (f^{-1})^n$ are used, then this equation is valid for all integers $n$.

The periodic procedure is applicable to pairs of functions simultaneously conjugate to translations. A pair of functions $f,g$ are simultaneously conjugate to translations if a function $u$ can be found that simultaneously satisfies both conjugacy equations, i.e., if for a single increasing 1 - 1 function $u$ and a pair of constants $k_f, k_g$

$$\text{(II.1)} \qquad \begin{cases} f(\cdot) = u^{-1}[u(\cdot) + k_f] \quad , \\ g(\cdot) = u^{-1}[u(\cdot) + k_g] \quad . \end{cases}$$

Note that simultaneously conjugate functions commute, i.e., if $f,g$ satisfy (II.1) then $fg(x) = u^{-1}[u(x) + k_f + k_g] = gf(x)$ and

$$\text{(II.2)} \qquad fg = gf \quad .$$

## II.2 Computing Ratios from Graphs

The periodic procedure is a means for computing the ratio $B = k_f/k_g$ for a pair of simultaneous conjugates satisfying (II.1). It can be used when the graphs of the conjugates are given but when all that is known about $u$ is that it is a strictly increasing, continuous real function.

To reintroduce the logic of the periodic procedure consider a pair of conjugates $f,g$ satisfying (II.1). For simplicity, we assume $k_f$ and $k_g$ are both positive. Since $k_{f^{-1}} = -k_f$, this results in no loss of generality.

In the next few paragraphs it will be shown how $k_f/k_g$ could be computed by using only the graphs of $f$ and $g$.

Suppose for some $x$, we have

$$f(x) \leq g(x) \; .$$

From (II.1) and the fact that $u^{-1}$ is an increasing 1 - 1 function, it follows that $f(x) \leq g(x)$ is equivalent to $k_f \leq k_g$. Consequently, if $f(x) \leq g(x)$ for any $x$, $k_f \leq k_g$ and $f(y) \leq g(y)$ for all $y$. Thus a superficial inspection of the graphs of $f$ and $g$ tells whether $B \leq 1$.

To get a more precise estimate of $B$, consider the iterates $f^n$ and $g^m$. Since $f^n(x)$ equals $f[f^{n-1}(x)]$ and $g^m = gg^{m-1}$, points on the graphs of iterates of $f$ and $g$ can be computed recursively from the graphs of $f$ and $g$, without referring to the function $u$ or the constants $k_f$ and $k_g$ of the conjugacy equations. Thus the graphs of $f$ and $g$ contain all the information needed to decide whether or not any inequality of the form

$$\text{(II.3)} \qquad f^n(x) \leq g^m(x)$$

is true. But since $f^n(\cdot) = u^{-1}[u(\cdot) + nk_f]$ and $g^m(\cdot) = u^{-1}[u(\cdot) + mk_g]$ are also simultaneously conjugate to translations, (II.3) is equivalent to $nk_f \leq mk_g$.

Thus, by referring only to the graphs of $f$ and $g$ for each integer $n$ we can find the smallest integer $m = m_n$ satisfying (II.3). Thus for $m = m_n$ we have the inequalities

$$\begin{cases} f^n(x) \leq g^m(x) & , \\ g^{m-1}(x) < f^n(x) & . \end{cases}$$

These inequalities are equivalent to

$$(m - 1)k_g < nk_f \leq mk_g$$

or

$$\text{(II.4)} \quad 0 \leq m_n/n - k_f/k_g < 1/n \quad .$$

Consequently, we can, in principle, use the graphs of f and g to decide whether inequalities of form $f^n(x) \leq g^m(x)$ are valid for large values of n and m and thereby approximate B with any desired accuracy. In this sense, the graphs of f,g contain all the information needed to compute the ratio.

## II.3 Computing Ratios from Segments of Graphs

Comparing iterates or, more generally, products in an ordered semi-group, plays a fundamental role in the logic of physical measurement (Levine, 1975). But applications to psychological data have been blocked by various difficulties. The most serious difficulty is the fact that the graphs of f and g are only known over a finite interval of the x axis, often a fairly narrow one. Consequently the higher iterates cannot be obtained, since the calculation of $f^n(x)$ requires that we evaluate f at x, at $x_1 = f(x)$, at $x_2 = f(x_1)$, etc., and we rapidly leave the domain where f is defined. Stated in another way, the domain of definition of $f^n$ rapidly gets smaller as n increases, and soon vanishes entirely. We have only rarely found empirical graphs of conjugates that could be used to directly test an inequality of the form $f^n(x) \leq g^m(x)$ for n or m greater than 5.

A related difficulty is that the graphs are not known exactly, but only approximated by finite tables of numbers having limited precision. Each composition reduces the precision, and again the comparison of higher iterates is difficult.

The periodic procedure uses an elementary functional equation and the properties of periodic functions to provide an indirect way for testing inequalities. In so doing it avoids the problem with higher iterates. At the same time it leads to natural ways to combine observations and overcome the difficulties with finite approximations. It also suggests ways to judge the degree to which data satisfies a model implying the conjugacy equations.

The periodic procedure begins with Abel's functional equation

$$(II.5) \qquad g(\cdot) = w^{-1}[w(\cdot) + 1]$$

where $g$ is a given function and $w$ is to be determined. Fortunately, it is easy to solve this equation for $w$. We begin the periodic procedure by computing a continuous, strictly increasing function $w$ that satisfies $g(x) = w^{-1}[w(x) + 1]$ for all values of $x$ in the domain of $g$. (References and a discussion of this computation are given in Section VI.1.)

This function is used to compute a function $\phi$ defined by

$$(II.6) \qquad \phi(x) = wfw^{-1}(x) \quad .$$

Since $\phi^n = (wfw^{-1})^n = wf^n w^{-1}$ and $g^m(x) = w^{-1}[w(x) + m]$, it follows that (II.3) $f^n(x) \leq g^m(x)$ is equivalent to

(II.7) $\phi^n(x) \leq x + m$ .

Thus if the higher iterates of $\phi$ could be computed, $k_f/k_g$ could be computed.

By using periodic functions, all the iterates of $\phi$ can be computed, even when the higher iterates of $f$ and $g$ cannot.

The condition (II.2) $fg = gf$ leads to periodic functions as follows. Let $p$ be the translation $p(x) = x + 1$. Then $g = w^{-1}pw$, and $fg = gf$ can be written $f(w^{-1}pw) = (w^{-1}pw)f$. Equivalently, $w[f(w^{-1}pw)]w^{-1} = w[f(w^{-1}pw)]w^{-1}$, that is,

$$\phi p = (wfw^{-1})p = p(wfw^{-1}) = p\phi$$

and

$$\phi(x + 1) = \phi(x) + 1 .$$

Thus if $\theta(x) = \phi(x) - x$, then

$$\theta(x + 1) = \phi(x + 1) - (x + 1)$$

$$= \phi(x) + 1 - (x + 1)$$

$$= \phi(x) - x$$

$$= \theta(x) ,$$

and $\Theta$ is periodic with period equal to one. Consequently, if $\Theta$ is defined on an interval of length equal to or greater than one, then $\Theta(x)$ can be regarded as known for all $x$. But in this case, $\phi(x)$ is also known for all $x$. For $\phi(x)$ is simply $\Theta(x) + x$. Consequently $\phi^n(x) = \phi[\phi^{n-1}(x)]$ can be computed recursively for all $n$ and $k_f/k_g$ can be computed with any desired accuracy.

In all the applications we know about, only points on short segments of graphs can be adequately estimated from data. (A segment of a graph of a function is simply the graph of the restriction of the function to an interval.) It is not possible to compute graphs of higher iterates of $f$ and $g$ from segments. However a segment of the graph of $g$ permits the computation of a segment of graph of $w$; and with a segment of the graph of $w$ and of $f$ we can compute a segment of the graph of $\phi$ and $\Theta$. But from a long enough segment of the graph $\Theta$ we can compute all of the graph of $\Theta$ and $\phi$. This permits us to check $\phi^n(x) \leq x + m$ and (II.3) for all $n,m$. Thus we can compute $B$ with any desired accuracy from segments too short to define higher iterates of $f$ and $g$.

This completes the review of the periodic procedure.

In the next two sections we prepare to apply the periodic procedure by showing that current psychometric theory implies that certain empirical curves are close to the graphs of simultaneous conjugates. The empirical curves permit us to compute a segment of the graph of a function such that if the psychometric theory were correct in every detail and if our data were based on enough observations, the graph of the empirical function would be indistinguishable from the graph of a periodic function. We will later use this fact to obtain a rough index of the appropriateness of the periodic procedure and a method for combining observations to increase measurement accuracy. In the results section we will report our application of the periodic procedure to psychometric data. It will be shown that the periodic procedure "works" in the sense that it agrees with a thoroughly tested alternative measurement procedure.

## III. SOME BASIC PSYCHOMETRICS

According to item characteristic curve theory and the logistic model, Lord's LOGIST parameter estimates and periodic procedure estimates should agree. In this section basic item characteristic curve theory, the logistic model and the logic of the LOGIST programs are reviewed. A generalization of the logistic model is then introduced, and some empirical results supporting the logistic model and LOGIST are cited.

### III.1 Item Characteristic Curve Theory and the LOGIST Programs

Item characteristic curve theory provides stochastic models for aptitude tests. The theory is designed for tests with many separate multiple choice items. Each candidate or test taker is assumed to have some (unknown) ability. His answers on each item are scored zero for wrong and one for right. The theory relates item scores to ability.

According to the theory the conditional probability of a randomly chosen candidate with ability $\theta$ correctly answering the $i$ -th item of a test is an increasing function of $\theta$ , $P_i(\theta)$ , called the item characteristic curve. The candidate's right and wrong answers are regarded as the outcome of a two-stage experiment. First an individual with some (unobserved) ability $\theta$ is sampled. Then a sequence of independent dichotomous random variables corresponding to the items, is observed. The probability that the $i$ -th item is scored correct is $P_i(\theta)$ .

Let $< e_1, e_2, e_3 \cdots e_n >$ be a vector of zeros and ones. The conditional probability that a randomly selected candidate has this particular pattern of correct and incorrect answers is

$$\text{(III.1)} \quad \prod_{i=1}^{n} P_i(\theta)^{e_i}[1 - P_i(\theta)]^{1-e_i} \quad .$$

We will later need to assume that the distribution of ability in the population of examinees has a continuous density. If the ability density is denoted by $f$ then the unconditional probability of observing the pattern of correct and incorrect answers is

$$\text{(III.2)} \quad \int f(\theta) \prod_{i=1}^{n} P_i(\theta)^{e_i}[1 - P_i(\theta)]^{1-e_i} \, d\theta$$

and the joint probability of sampling an examinee with ability $\theta$ in an interval $T$ and the given pattern of answers is

$$\text{(III.3)} \quad \int_T f(\theta) \prod_{i=1}^{n} P_i(\theta)^{e_i}[1 - P_i(\theta)]^{1-e_i} \, d\theta \quad .$$

The goal of testing is to estimate the individual examinee's ability from his pattern of right and wrong answers. If the item characteristic curves $P_i$ are specified, this is fairly routine. In this paper we are primarily concerned with using data to specify item characteristic curves.

The LOGIST programs compute maximum likelihood estimates of the item characteristic curves. Each $P_i$ is assumed to have the form

$$P_i(t) = c_i + (1 - c_i)L[a_i(t - b_i)]$$

where L is the logistic function, $L(t) = 1/[1 + e^{-t}]$, and $a_i$, $b_i$, $c_i$ ~~are real~~ parameters controlling the shape of $P_i$. (The role of these parameters and the logistic function is discussed in the next section.) Ordinarily a set {a} of at least a thousand examinees is processed by LOGIST. The typical candidate a produces a vector of item scores which has conditional probability

$$\prod_{i=1}^{n} P_i(\Theta_a)^{e_{ia}}[1 - P_i(\Theta_a)]^{1-e_{ia}} .$$

LOGIST forms the product

$$\text{(III.4)} \quad \prod_{a} \prod_{i=1}^{n} P_i(\Theta_a)^{e_{ia}} Q_i(\Theta_c)^{1-e_{ia}}$$

and iteratively seeks a vector of abilities $\hat{\Theta}_a$ and item parameters $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$ maximizing (III.4).

The LOGIST programs have many options and are regularly revised. The data and LOGIST application used for reference in this paper have been described by Lord (1968). We will review special features of that study as needed.

III.2 The Form of Item Characteristic Curves

Various item characteristic curve models differ in their assumptions about the shape and functional form of the $P_i$'s. There is considerable evidence to support the logistic and normal models. In this section we review these models and introduce a generalization called the linear model.

The logistic model assumes that each $P_i$ has functional form

$$(III.5) \quad P_i(\theta) = c_i + (1 - c_i)L[a_i(\theta - b_i)]$$

where L is the S-shaped logistic function. A plot of the graph of $P_i$ appears S-shaped with a left asymptote of $c_i$ and a right asymptote of 1. $P_i$ is 1 - 1, strictly increasing, continuous and has a single point of inflection.

The parameters $a_i$, $b_i$, $c_i$ describe three more or less independent properties of items. An increase in the "difficulty" parameter $b_i$ shifts the graph of $P_i$ along the x-axis and decreases the probability of a correct answer at all levels. Variation of the positive "discrimination" parameter most conspicuously affects the steepness of $P_i$ at its point of inflection. Geometrically it stretches or contracts the graph. The "guessing" parameter $c_i$ controls the height of the left asymptote of $P_i$. This parameter is sometimes interpreted as the limiting probability of an examinee with no preparation or ability correctly guessing the answer to the item.

A familiar model commonly associated with Rasch (1960) postulates

$$P_i(\theta) = \theta/(\theta + b_i)$$

where ability is assumed to be positive. Birnbaum (1968, page 402) has pointed out that this model can be regarded as a special case of the logistic model since

$$\Theta/(\Theta + b_i) = c_i + (1 - c_i)L[a_i(\Theta^*_\circ - b^*_i)] \quad ,$$

where $c_i = 0$ , $a_i = 1$ , $\Theta^* = \log \Theta$ , $b^*_i = \log b_i$ . We will use this model only to motivate some of the definitions of the next section. Unfortunately, it is not sufficiently general to describe the data we wish to study.

By replacing the logistic function $L(x)$ by the normal integral $\Phi(x)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

one obtains the normal model. Since $|\Phi(x) - L(Dx)|$ where $D$ is a known constant is very small for all values of $x$ , there is very little difference between the shapes of icc's for the normal and logistic model.

In this paper we assume that tests satisfy a generalization of the logistic and normal models called the <u>linear model</u>. A set of continuous, strictly increasing icc's $\{P_i\}$ satisfy a linear model if there exists a continuous, strictly increasing probability distribution $P$ and real constants $\{a_i\}$ , $\{b_i\}$ , $\{c_i\}$ such that for all $\Theta$

$$P_i(\Theta) = c_i + (1 - c_i)P[a_i(\Theta - b_i)] \quad .$$

(Such icc's are said to satisfy a "linear" model because any two curves can be equated by a linear or affine transformation of the plane having form $< x,y > \rightarrow < ax + b, cy + 1 - c >$.)

For concreteness, P may be thought of as either the logistic or normal. But the analysis presented in this paper is equally applicable to the logistic, normal or other continuous, strictly increasing distribution function.

The fact that the periodic procedure does not require complete specification the functional form of the $P_i$ has important practical implications. It is closely related to the invariance of Section IV.2, which greatly simplifies data analysis.

Our goal is to evaluate the periodic procedure by using it to estimate item parameters from some frequently analyzed data, evidently well described by item characteristic theory. Lord (1970) has already obtained impressive evidence for the adequacy of logistic model. The close agreement between parameter estimates (obtained without the logistic assumption) and LOGIST parameter estimates provides further support for the logistic model.

## IV. ITEM-ITEM CURVES

The following subsections are used to introduce and discuss the basic properties of the item-item ability curve, a generally handy adjunct to item characteristic curve theory and an essential tool for the application of periodic functions. To motivate this new development, recall that a continuous 1 - 1 transformation of ability was shown to change the item characteristic curves of the Rasch model into the curves of a one-parameter logistic model. Item-item curves will be seen to carry very much the same information as icc's. But we will show (Section IV.2) that they are invariant under all 1 - 1 continuous transformations of ability. By using this invariance we are able to specify consistent estimators of points on the item-item ability curves (Section IV.3). Item-item ability curves can be related to conjugates of translations in various ways. We outline the way actually used (Section IV.4) in this study along with alternative ways, which may be preferred for free-response data and other psychometric data for which the periodic procedure may be applied.

### IV.1 Item-Item Curves and Graphs

By considering pairs of items, one defines item-item ability curves. The (ij) -th item-item ability curve is the subset of the unit square.

$$\{ \langle x,y \rangle : \text{for some ability } \theta, \; x = P_i(\theta) \text{ and } y = P_j(\theta)\} .$$

Every item characteristic curve considered in this paper is 1 - 1 and thus has an inverse. Thus for each $P_i$ and every $t$ in the range of $P_i$ we may write $t = P_i[P_i^{-1}(t)]$ . Consequently, each item-item ability curve can be written in the form

$$\{ < t, P_j[P_i^{-1}(t)] > : \text{for some ability } \theta , \ t = P_i(\theta)\} \ .$$

In other words, $(ij)$ -th item-item ability curve is simply the graph of the function $P_j P_i^{-1}$ .

IV.2 Item-Item Curves and Transformations of Ability

Since ability $\theta$ is not observed, one is led to consider various observed indices of ability such as proportion of correct items, formula scores, grade point average, composite scores on a battery of tests not containing the items being studied, maximum likelihood estimate of ability, etc. If $\xi$ is such an index, then $(i,j)$ -th item-item index curve of a test is the curve

$\{ < x,y > :$ for some value of $\xi$ , $x$ is the conditional probability of correctly answering the $i$ -th item and $y$ , correctly answering the $j$ -th$\}$ .

There is a special case, which is important for theoretical developments and which is conceptually important for our application. Suppose an index $\xi$ is a 1 - 1 function of ability $\theta$ , i.e., that for some invertible function $v$ we have $\xi = v(\theta)$ . In this case the item-item index curves and item-item ability curves are equal. This is true because for each value $\xi_o$ the conditional probability

$$\text{Prob}\{\text{item } i \text{ is correct} | \text{index } \xi = \xi_o\}$$

equals

$$\text{Prob}\{\text{item } i \text{ is correct} | \text{ability } \theta \text{ equals } v^{-1}(\xi_o)\}$$

which equals $P_i[v^{-1}(\xi_o)]$. Thus the item-item index curve is the graph of the function $(P_j v^{-1})(P_i v^{-1})^{-1}$ which is clearly equal to the graph of the function $P_j P_i^{-1}$. It follows that the $(ij)$-th item-item index curve equals the $(ij)$-th item-item ability curve.

This argument shows that the same item-item curves are obtained with any 1 - 1 transformation of ability. In particular, strictly increasing, continuous transformations of ability leave item-item ability curves invariant.

This simple invariance turns out to be very important. It greatly simplifies the estimation of item-item curves. It makes possible results (presented in the next section) that are used to relate observed item-item proportion correct curves and item-item ability curves. It also implies, incidentally, that two models such as the Rasch and one-parameter logistic with different item characteristic curves can have exactly the same item-item curves.

## IV.3 Relating Item-Item Curves to Observations

The purpose of this section is to describe some asymptotic results that are used to estimate points on item-item ability curves.

Consider repeatedly sampling examinees from a population in which ability has a normal or some similar smooth probability density function $f$. Each sampled examinee takes the same long test. As noted in Section III.1 both ability $\theta$ and the item scores $u_i$ can be regarded as random variables. The cited asymptotic results follow from regularity conditions and from formula (III.3), the formula for the joint distribution of ability and the first $n$ scores.

We intend to comment on the suitability of the observed proportion of correct answers as an index of ability. Thus we must define the proportion correct random variable $z_n$

$$z_n = \frac{1}{n} \sum_{i=1}^{n} u_i$$

i.e., the random variable giving the average item score or the proportion of correct answers on the first $n$ items.

Let $\zeta$ be a fixed proportion in the range of each $P_i$. Consider the average $\mu_n$ of the first $n$ item characteristic curves

$$\mu_n(t) = \frac{1}{n} \sum_{i=1}^{n} P_i(t) \quad .$$

Since according to the linear model each $P_i$ is strictly increasing and continuous, the average $\mu_n$ must be also. Thus $\mu_n$ is 1 - 1 and has a strictly increasing, continuous inverse $\xi_n$ . Since $\zeta$ is in the range of each $P_i$ , $\zeta$ is the range of $\mu_n$ and $\xi_n(\zeta)$ is defined.

As we consider longer and longer initial segments of the sequence of test items, the sequence of numbers $\{\xi_n(\zeta)\}$ may or may not converge. However, it still can be shown that under quite general conditions it is possible to specify a sequence of "proportion interval widths" $s_n$ that slowly decrease to zero such that for each item i the difference

$$\mathcal{E}\{u_i \mid |z_n - \zeta| \leq s_n\} - P_i[\xi_n(\zeta)]$$

tends to zero as n increases.

In other words, asymptotically, the expected item score for item i --given that proportion correct on the first n items is close to $\zeta$ --is approximately equal to the i -th item characteristic curve composed with a strictly increasing continuous transformation relating the set of possible proportions to the ability continuum. Furthermore, the transformation $\xi_n$ is independent of i .

To apply this fact, one considers a finite set of proportions

$$\zeta_1, \zeta_2, \ldots, \zeta_R$$

and a pair of items, say for definiteness items one and two. A very large sample A of independently sampled examinees is selected and administered an n item test. For each candidate $a \in A$ , proportion correct $z_n(a)$ is computed and used as an index to define R subsamples

$$A_r = \{a \in A: |z_n(a) - \zeta_r| \leq s_n\} \quad , \quad r = 1,2,\ldots,R \quad .$$

As the size $N(A_r)$ of these subsamples becomes large, each observed proportion $\rho_{ir}$

$$\rho_{ir} = \frac{N\{a \in A_r | a \text{ answers item } i \text{ correctly}\}}{N(A_r)}$$

will, with high probability, be close to its expected value

$$\mathcal{E}(u_i \big| \, |z_n - \zeta_r| \leq s_n) \quad ,$$

and therefore close to $P_i[\xi_n(\zeta_r)]$ . Thus the points in the plane $< \rho_{1,r}, \rho_{2,r} >$ are likely to be close to the points $< P_1\xi_n(\zeta_r), P_2\xi_n(\zeta_r) >$ on the item-item ability curve $< P_1, P_2 >$ . This follows from the fact that each $\xi_n$ is 1 - 1 and the invariance described in the preceding section.

Proofs of these results and a discussion of rates of convergence will be discussed in a separate paper currently being prepared.

In this study empirical item-item proportion correct curves have been used as estimates of points on item-item ability curves.

IV.4. Relating Item-Item Curves to Conjugates

Before the periodic procedure can be applied, estimates of points on conjugates of translations must be obtained. We have just shown that empirical item-item proportion correct curves are estimates of points on theoretical item-item ability curves. In this section it is shown that the theoretical iia curves can be related to conjugates in various ways. Each way suggests a somewhat different data analysis procedure. The special purpose of this study and the special nature of our data have selected a particular way, presented below. Alternative ways are noted in passing.

To begin, consider a one parameter model version of the linear model in which $c_i = 0$ and $a_i = 1$, so that every icc has form

$$P_i(\Theta) = P(\Theta - b_i) \quad .$$

The $(i,j)$ -th item-item ability curve is simply the graph of the conjugate

$$x \to P\ell_{ij}P^{-1}$$

where $\ell_{ij}$ is the translation $\Theta \to \Theta + b_i - b_j$ .

In the general linear model each icc has form

$$P_i(\Theta) = c_i + (1 - c_i)P\ell_i(\Theta)$$

where $c_i$ is the "guessing parameter," and $\ell_i$ is the linear transformation

$$\Theta \to a_i(\Theta - b_i) \doteq \ell_i(\Theta) \quad .$$

The item-item ability curves here are the graphs of functions of form

$$C_jP\ell_j(C_iP\ell_i)^{-1}$$

where each function $C_i$ is defined by $C_i(t) = c_i + (1 - c_i)t$. This function can be rewritten as $(C_jP)(\ell_j\ell_i^{-1})(C_iP)^{-1}$ where $\ell_j\ell_i^{-1}$ can easily be shown to be

$$\theta \to \frac{a_j}{a_i}\theta + a_j(b_i - b_j) \quad .$$

Thus if item $i$ is compared with an item $j$ with the same discrimination parameter, $a_j/a_i = 1$ and $(i,j)$ -th item-item ability curve is the graph of a function very much like the graph of a conjugate,

$$C_jP(\ell_j\ell_i^{-1})(C_iP)^{-1} \quad .$$

If $c_i$ is assumed to be the reciprocal of the number of multiple choice alternatives and both items have the same number of alternatives then $C_j = C_i$ and

$$C_jP(\ell_j\ell_i^{-1})(PC_i)^{-1} = C_jP(\ell_j\ell_i^{-1})(PC_j)^{-1}$$

is again a conjugate of a translation.

When the $c_i$ 's are set equal to the reciprocal of the number of choices which can otherwise be regarded as known, a simple linear transformation of the plane can be specified and used to remove the effects of the $c_i$ 's. Using elementary algebra, one can verify that the transformation

$$< x,y > \to < (x - c_i)/(1 - c_i),\ (y - c_j)/(1 - c_j) >$$

carries the $(ij)$ -th item-item ability curve onto the graph of the function $P(\ell_j\ell_i^{-1})P^{-1}$.

In this study it is appropriate to regard both the $a_i$ 's and $c_i$ 's as known. (A discussion of this point follows in the next section.) Consequently we were able to pair items with equal discrimination and rescale empirical item-item proportion correct curves to obtain estimates of points on the graphs of conjugates of translations.

IV.5 Concluding Comments on Our Method of Using Item-Item Curves

In this section we wish to state the rationale for our way of estimating points on the graph of conjugates.

The preceding sections show what would happen in a large scale simulation of item characteristic curve theory with, say, one parameter logistic icc's, a bounded sequence of item difficulties and a normal distribution of abilities: If a pair of conjugates were approximated from empirical item-item proportion correct curves, smoothed and used as input for the periodic procedure, then as test length and examinee sample size increased, the periodic procedure would almost surely give essentially perfect estimates of item difficulties. Some nontrivial mathematical questions would have to be worked out to properly qualify and prove this assertion in detail. But we do not develop this line of reasoning any further here because it is tangential to the previously stated goals of this study.

The central question for this research is not mathematical; it can only be answered by data. It is: Can the periodic procedure be useful with current tests, which are only imperfectly described by item characteristic curve theory?

To make our point clear, recall that item characteristic curve theory is an idealization, i.e., an abstract, incomplete description of aptitude test data. It lacks, for example, an explicit account of the complex interdependences between items referring to the same reading passage on the typical verbal aptitude test.

However, icc has proven to be a powerful tool for investigating important problems. Lord (1975) cites 17 recently published applications. Icc's role in studies in progress as well as the absence of a likely alternative makes it fairly certain that icc will be even more widely used in the next decade.

We wish to consider the type of data that is going to continue to be analyzed with icc. By reanalyzing such data with the periodic procedure we intend to decide whether the periodic procedure estimates of item difficulty agree sufficiently well with standard procedures to justify further development.

The reference estimation procedure, Lord's LOGIST program, simultaneously fit $a_i$ 's, $b_i$ 's, and $c_i$ 's (as well as $\theta$ 's) for samples of examinees to the logistic model. In its present form, the periodic procedure estimates only $b_i$ 's. Since our goal is to compare periodic procedure estimates with LOGIST $b_i$ 's and not to document a new estimation scheme, it is legitimate and desirable to use LOGIST estimates of $\hat{a}_i$ 's and $c_i$ 's. Thus we are led to the following steps in estimating points on graphs of conjugates:

1. pair items with equal LOGIST $a_i$ 's.
2. compute empirical item-item proportion correct curves for these item pairs.
3. use LOGIST $c_i$ 's to rescale and obtain estimates of points on the graphs of conjugates of translations.

All of the essential details of the implementation of these steps are presented in subsequent sections.

## V. DATA AND RESULTS

Our major finding is close in agreement between LOGIST and periodic estimates. In this section we summarize the basic facts about the data and computations; additional details are in the references and the sections on computation.

Lord (1968) applied LOGIST to a sample of 2,862 SAT-V candidates. Number right scores on the SAT-M were used to obtain a sample with a relatively large proportion of low ability candidates. The $c_i$ guessing parameters were estimated to be equal to the observed proportions correct for low scoring candidates. Omitted and not reached items were scored as if they had been answered incorrectly. The $a$'s, $b$'s, and $\theta$'s were estimated by minimizing equation III.4.

In the following $\hat{x}$ and $\tilde{x}$ are used to denote LOGIST and periodic estimates of a parameter $x$.

LOGIST estimates, provided by Lord, were used to select a subtest from the SAT-V. Means, standard deviations and other summary statistics are given in Table 1. The LOGIST estimated parameters of the seven item subtest are given in Table 2. These are all the items with $\hat{a}_i$ (LOGIST estimated $a_i$) equal to $1.22 \pm .02$. This particular constant was chosen because many items happened to have $\hat{a}_i$'s close to 1.22.

Item-item proportion correct curves were estimated by first computing item test regressions from the entire administration of 103,275 candidates. The item test regression for item $i$ is essentially the vector proportions ( $\rho_{ir}$ ) of Section III.3 where $r$ is the index of the proportion $\zeta_r$.

Table 1

Summary Statistics for LOGIST Estimated Parameters

| | $\hat{a}_i$ | $\hat{b}_i$ | $\hat{c}_i$ |
|---|---|---|---|
| Mean | 1.09 | .65 | .15 |
| Standard Deviation | .385 | .855 | .05 |
| Median | 1.07 | .77 | .16 |
| Maximum | 2.03 | 2.425 | .2 |
| Minimum | .40 | -1.52 | .04 |

Table 2

LOGIST Estimates for Subtest

| item number | $\hat{a}_i$ | $\hat{b}_i$ | $\hat{c}_i$ |
|---|---|---|---|
| 51 | 1.2058 | -.5404 | .20 |
| 6 | 1.2072 | .7198 | .12 |
| 77 | 1.2133 | .7026 | .14 |
| 20 | 1.2170 | 1.6270 | .09 |
| 80 | 1.2180 | 1.0919 | .15 |
| 14 | 1.2321 | .5056 | .20 |
| 59 | 1.2341 | -.6478 | .20 |

More explicitly,

$$\rho_{ir} = N_{ir}/M_{ir}$$

where

$N_{ir}$ = the number of candidates answering item $i$ correctly and either $2r$ or $2r - 1$ other items correctly;

$M_{ir} = N_{ir}$ + the number of candidates answering item $i$ incorrectly and either $2r$ or $2r - 1$ other items correctly.

This minor departure from III.3 avoids the spurious dependencies discussed in Lord and Novick (1968, Section 16.4.1) by conditioning on the number correct on the whole test, except for item $i$, rather than on the whole test.

The plot of the item test regression (i.e., the plot of $\rho_{ir}$ as a function of $\zeta_r$) generally appears to be a set of closely spaced points on the graph of a smooth, monotonic function. This is especially true for intermediate values of $\zeta_r$ for which the observed proportions are based on a large number of examinees. Some nonmonotonicities were observed for very low and very high proportions correct. This is to be expected since there were few examinees with very high or very low proportion correct scores. (In Section IV.2 we report a technique for preventing these unstable proportions from affecting our item difficulty estimates.)

As a first step towards obtaining curves suitable for the periodic procedure, for each item $i$ a vector ( $\rho^*_{ir}$ ) was computed. The first entry $\rho^*_{i1}$ is the asymptote $\hat{c}_i$ used by LOGIST. The last entry $\rho^*_{iR}$ is one. The other values are computed by initially setting $\rho^*_{ir} = \rho_{ir}$ and then applying the algorithm described by Kruskal (1964, Section 8) to compute the monotonic array of points beginning with $c_i$ and ending with one maximizing a quadratic index of goodness of fit. Ties introduced by the algorithm were broken by making very small adjustments of the equal $\rho^*_{ir}$ 's. In this way each item was used to generate a vector of strictly increasing numbers ( $\rho^*_{i,r}$ ) approximately equal to the observed proportions ( $\rho_{ir}$ ).

The $i$ -th and $j$ -th vectors were then used to specify estimates $\{ < \rho^*_{ir}, \rho^*_{jr} > \}$ of points on the $ij$ -th item-item proportion correct curve. The $c_i$ 's were then used as described in Section IV.4 to estimate points on conjugates.

In this way the $ij$ -th pair of subtest items was used to specify an array of points $\{ < x_i(r), y_j(r) > \}$ in the plane. Since both $x_i(r)$ and $y_j(r)$ are strictly increasing functions of $r$ , these points are on the graph of at least one strictly increasing, continuous real function. By linearly interpolating between points we selected one such function $f_{ij}(\cdot)$ for each $ij$ .

According to the linear model $x_i(r)$ is an estimate of $u[\xi_n(\zeta_r) - b_i]$ and $y_j(r)$ is an estimate of $u[\xi_n(\zeta_r) - b_j]$ where $\xi_n$ is the 1 - 1 function of Section III.4, $b_i$ is the item difficulty and $u(t) = P[1.22t]$ for $P$ equal to the probability distribution of the linear model. Thus the points $< x_i(r), y_j(r) >$ are estimates of $< u[\xi_n(\zeta_r) - b_i], u[\xi_n(\zeta_r) - b_j] >$ on the graph of the function $u[u^{-1}(\cdot) + b_i - b_j]$ . Since the x 's and the y 's are closely spaced, the graph of $f_{ij}$ will be (pointwise) close to the graph of the conjugate $u[u^{-1}(\cdot) + b_i - b_j]$ , and we are ready to try the periodic procedure.

We selected three different functions $f_{k\ell}$ and solved Abel's equation for each of them. (Comments on the selection of the $f_{k\ell}$ follow in this and later sections.) For each $k\ell$ we used the same solution to Abel's equation to estimate

$$B(ij,k\ell) = (b_i - b_j)/(b_k - b_\ell)$$

for all pairs ij of items in the subtest.

The results are tabulated in Tables 3a, 3b, 3c and plotted in Figures 1a, 1b, 1c. Since our programs always estimate $B(ii,k\ell)$ to be zero and $B(ji,k\ell)$ to be $-B(ij,k\ell)$ only 21 of the possible 49 $\tilde{B}(\cdot\cdot,k\ell)$ need be given. Only the positive $\tilde{B}$ are tabulated and plotted. (The trivial estimate $\tilde{B}(k\ell,k\ell) = 1$ is tabulated but not plotted.)

The periodic estimates are in the second column. The order of the rows was determined so that $\hat{B}(ij,k\ell)$ (in the third column) increased from the first to the last row in each table. The columns labelled E and I will be used to illustrate a point considered in a later section.

The periodic estimates are very close to the LOGIST estimates. To make this obvious we computed an overall estimate of $b_i$ from each table by averaging $\tilde{B}(i\cdot,k\ell)$ . That is, for each $k\ell$ we computed

Table 3a

Estimates of B(ij;59,77) with Indices

| i | j | $\tilde{B}$ | $\hat{B}$ | E | I |
|---|---|---|---|---|---|
| 77 | 6 | .0009 | .013 | .870 | .66 |
| 59 | 51 | .082 | .080 | .012 | .16 |
| 14 | 77 | .138 | .146 | .028 | .34 |
| 14 | 6 | .125 | .159 | .120 | .15 |
| 6 | 80 | .253 | .276 | .043 | .058 |
| 77 | 80 | .305 | .288 | .029 | .039 |
| 80 | 20 | .417 | .396 | .026 | .018 |
| 14 | 80 | .405 | .434 | .035 | .014 |
| 6 | 20 | .654 | .672 | .014 | .021 |
| 77 | 20 | .709 | .685 | .017 | .020 |
| 51 | 14 | .739 | .775 | .024 | .023 |
| 14 | 20 | .846 | .830 | .0095 | .0088 |
| 59 | 14 | .811 | .854 | .026 | .013 |
| 51 | 77 | .9225 | .920 | .0014 | .0025 |
| 51 | 6 | .999 | .933 | .034 | .0023 |
| 59 | 77 | 1.000 | 1.000 | .000 | .00003 |
| 59 | 6 | 1.000 | 1.013 | .0065 | .0022 |
| 51 | 80 | 1.200 | 1.209 | .0037 | .0031 |
| 59 | 80 | 1.250 | 1.288 | .015 | .0043 |
| 51 | 20 | 1.552 | 1.605 | .017 | .0053 |
| 59 | 20 | 1.594 | 1.685 | .028 | .0091 |

Table 3b

Estimates of B(ij;51,77)

| i | j | $\tilde{B}$ | $\hat{B}$ | E | I |
|---|---|---|---|---|---|
| 77 | 6 | .024 | .0138 | .27 | .72 |
| 59 | 51 | .0875 | .0864 | .006 | .17 |
| 14 | 77 | .159 | .1585 | .016 | .36 |
| 14 | 6 | .138 | .1723 | .111 | .182 |
| 6 | 80 | .2805 | .2994 | .032 | .065 |
| 77 | 80 | .333 | .3132 | .031 | .025 |
| 80 | 20 | .4545 | .4305 | .027 | .026 |
| 14 | 80 | .434 | .4717 | .042 | .013 |
| 6 | 20 | .700 | .7298 | .021 | .02 |
| 77 | 20 | .763 | .7437 | .013 | .0077 |
| 51 | 14 | .772 | .8415 | .043 | .02 |
| 14 | 20 | .917 | .9022 | .0081 | .0098 |
| 59 | 14 | .857 | .9279 | .040 | .011 |
| 51 | 77 | 1.000 | 1.0000 | .000 | .00004 |
| 51 | 6 | 1.000 | 1.0138 | .007 | .0024 |
| 59 | 77 | 1.087 | 1.0864 | .0003 | .0017 |
| 59 | 6 | 1.001 | 1.1002 | .047 | .0029 |
| 51 | 80 | 1.333 | 1.3132 | .008 | .0029 |
| 59 | 80 | 1.333 | 1.3996 | .024 | .0044 |
| 51 | 20 | 1.727 | 1.7437 | .005 | .0017 |
| 59 | 20 | 1.800 | 1.8301 | .0083 | .003 |

Table 3c

Estimates of B(ij;77,20)

| i | j | $\tilde{B}$ | $\hat{B}$ | E | I |
|---|---|---|---|---|---|
| 77 | 6 | .0004 | .019 | .959 | .61 |
| 59 | 51 | .167 | .116 | .18 | .26 |
| 14 | 77 | .343 | .213 | .234 | .33 |
| 14 | 6 | .240 | .232 | .017 | .14 |
| 6 | 80 | .3335 | .403 | .094 | .043 |
| 77 | 80 | .391 | .421 | .037 | .037 |
| 80 | 20 | .750 | .579 | .13 | .12 |
| 14 | 80 | .500 | .634 | .12 | .029 |
| 6 | 20 | 1.000 | .981 | .0096 | .022 |
| 77 | 20 | 1.000 | 1.000 | .00 | .00004 |
| 51 | 14 | .857 | 1.132 | .14 | .028 |
| 14 | 20 | 1.333 | 1.213 | .047 | .015 |
| 59 | 14 | 1.000 | 1.248 | .11 | .010 |
| 51 | 77 | 1.356 | 1.345 | .004 | .010 |
| 51 | 6 | 1.214 | 1.363 | .058 | .0072 |
| 59 | 77 | 1.500 | 1.461 | .013 | .019 |
| 59 | 6 | 1.275 | 1.479 | .074 | .0083 |
| 51 | 80 | 1.500 | 1.766 | .082 | .030 |
| 59 | 80 | 1.5685 | 1.882 | .091 | .023 |
| 51 | 20 | 2.333 | 2.345 | .0025 | .0013 |
| 59 | 20 | 2.429 | 2.461 | .0065 | .0019 |

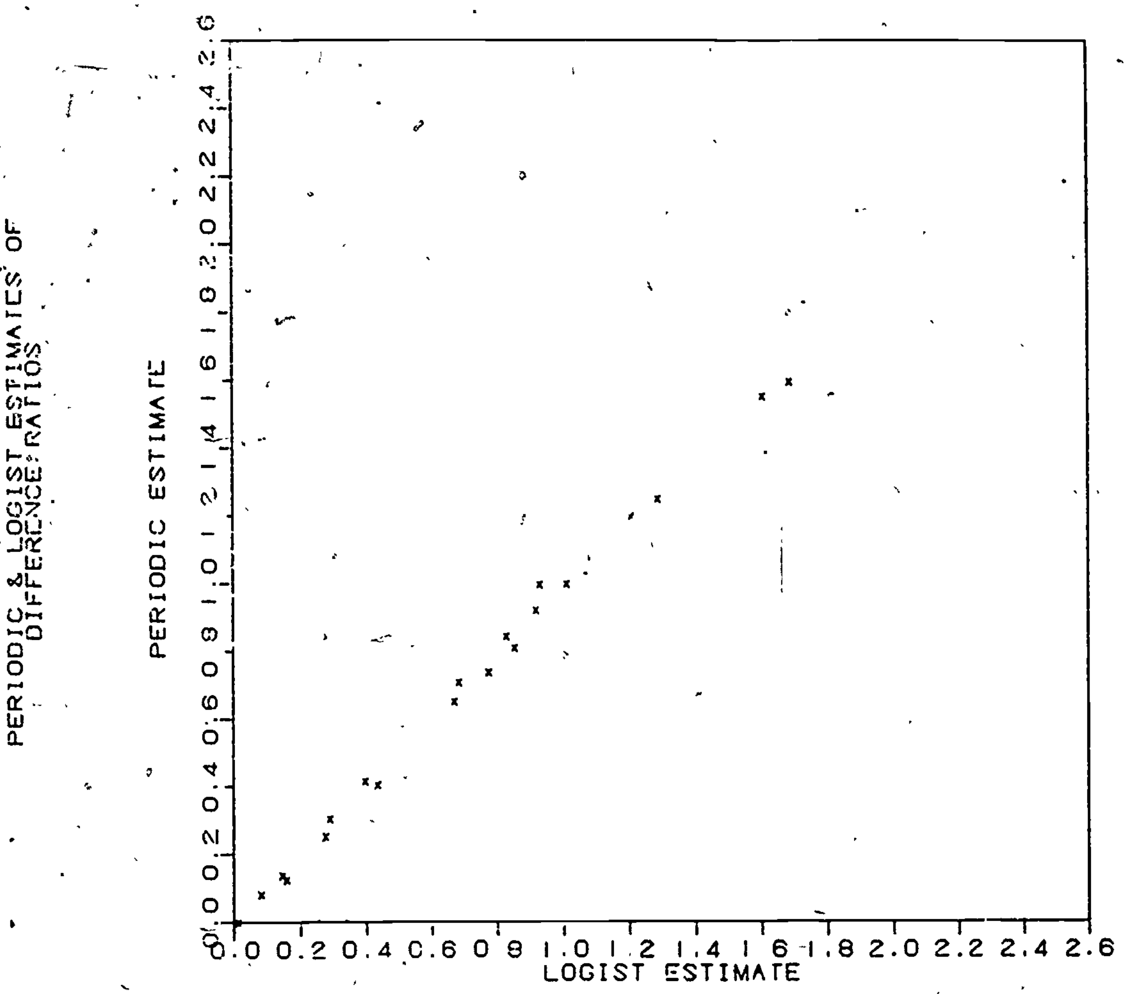


Figure 1a. Periodic and LOGIST estimates of B for k/ = 59,77 .

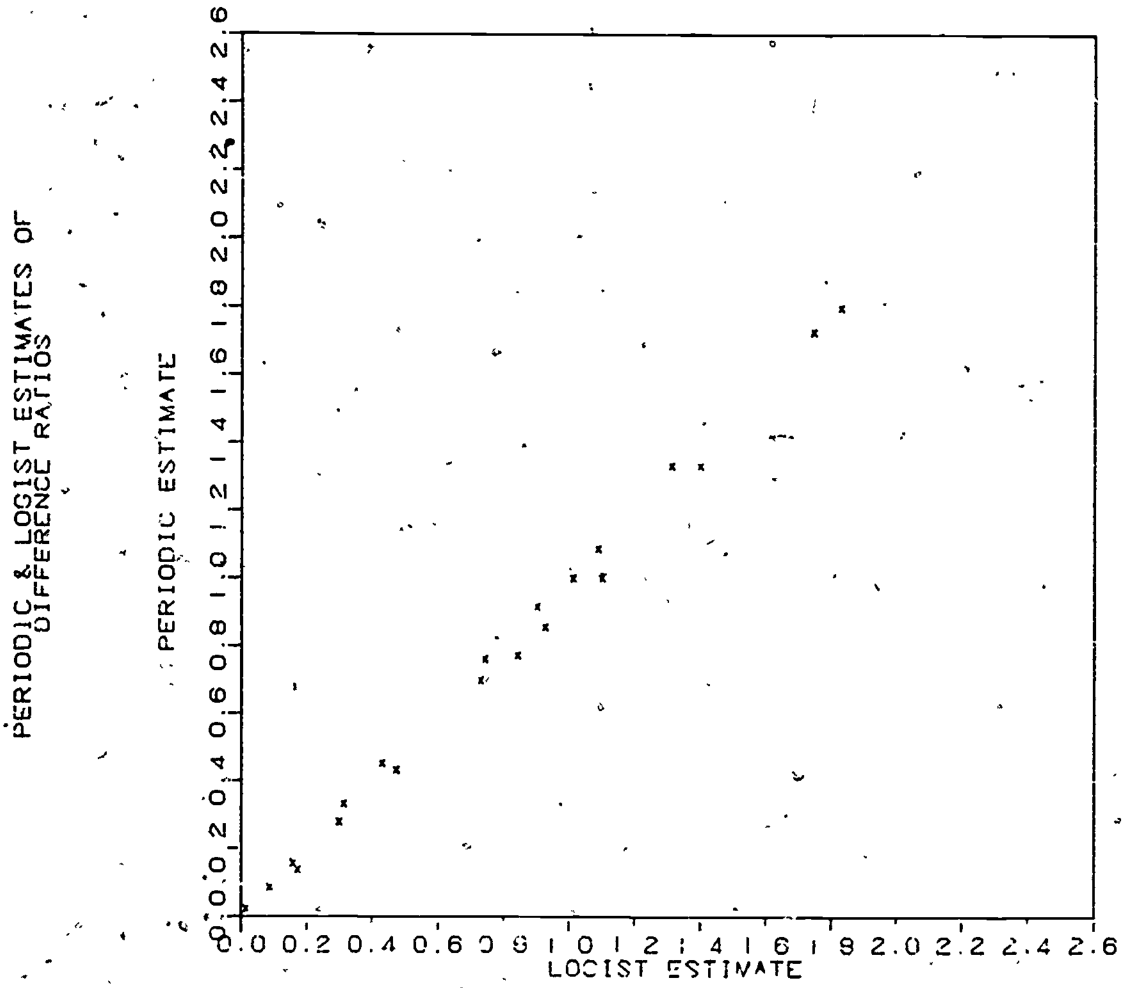


Figure 1b. Periodic and LOGIST estimates of B for k,ℓ = 51,77.

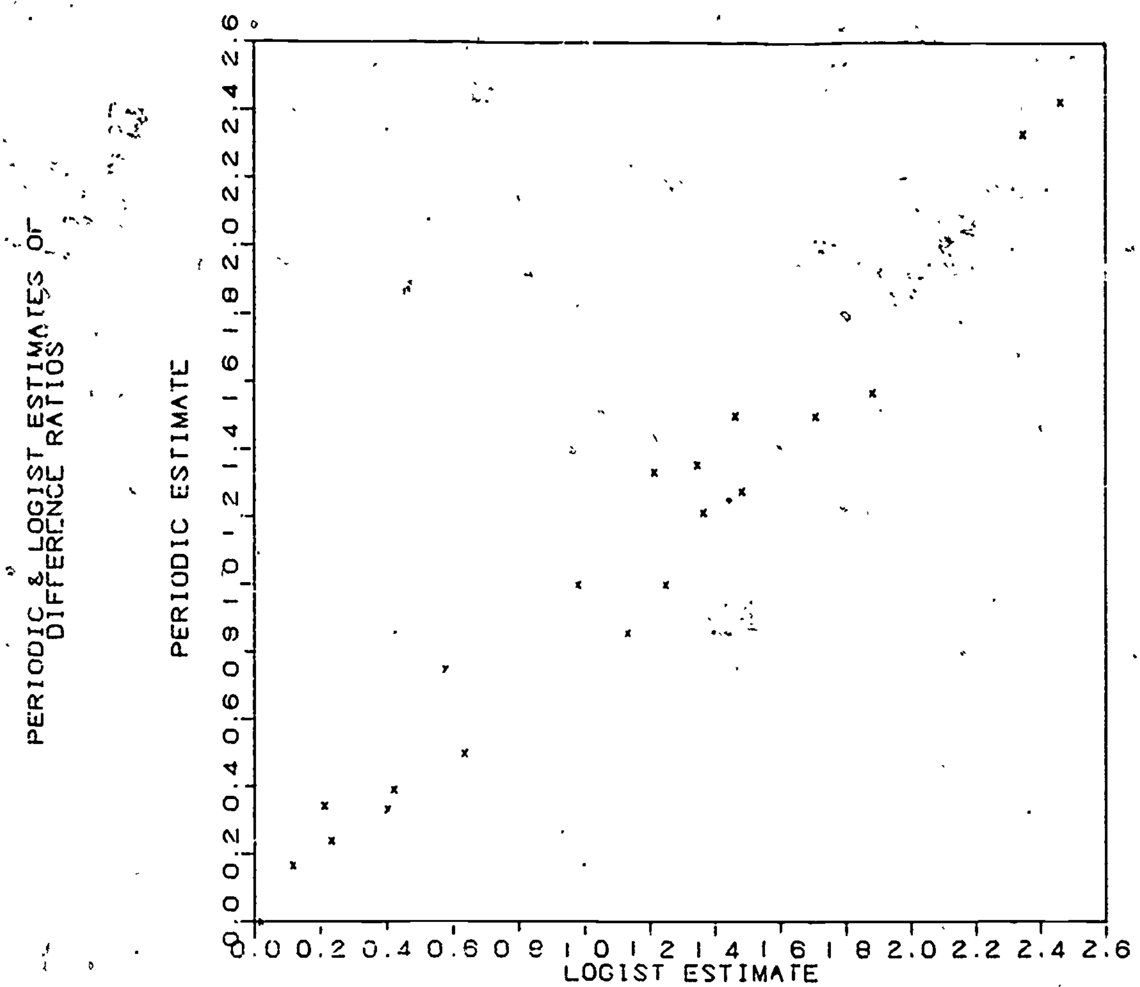


Figure 1c. Periodic and LOGIST estimates of B for $k, \ell = 77, 20$.

$$\tilde{b}_{i,k\ell} = \frac{1}{7} \sum_j \tilde{B}(ij,k\ell) \quad .$$

Since according to the linear model, the $b_i$ are on an interval scale, comparison of the $\hat{b}_i$ and $\tilde{b}_i$ can be simplified by rescaling the $\tilde{b}_{i,k\ell}$ so that they have the same mean and variance as the $\hat{b}_i$ . Our results are in Table 4.

Lord (1975) found the standard error of estimate for item difficulty (the square root of the mean squared error of prediction from the regression of $\hat{b}_i$ on $b_i$) to be .196 in a recent computer simulation of the SAT-V. The differences between the $\hat{b}_i$ and $\tilde{b}_i$ in Table 4 are so small relative to .196 that the periodic estimates and LOGIST estimates can be considered interchangeable.

We now reconsider the choice of the pair of items $k\ell$ . This choice is important because if the items are atypical or if $f_{k\ell}$ make poor use of the available data, then for every $ij$ the estimate of $B(ij,k\ell)$ will be adversely affected. A pair of items can make poor use of data in many ways. For example, if $b_k - b_\ell$ is very large, then virtually all of the data will be used to define a segment of $\theta$ less than unit length, and the periodic procedure will give poor estimates. If $b_k - b_\ell$ is nearly zero, then small sampling errors will have large effects on $\theta$ and again the periodic procedure will give poor estimates.

A straightforward, objective way to reject unacceptable $f_{k\ell}$ without referring to $b_k - b_\ell$ would be to compute periodic estimates for all pairs $i < j$ in the subtest and then test the hypothesis that $B(ij,k\ell)$ is equal to the difference $X_i - X_j$ between two (unknown, estimated) numbers plus a small random error. An alternative way is introduced in Section VI.3.

Table 4

LOGIST and Periodic Item Difficulty Estimates

| $i$ | $\hat{b}_i$ | $\tilde{b}_{i,k\ell}$ | | |
|---|---|---|---|---|
| | | $k\ell=59,77$ | $k\ell=51,77$ | $k\ell=77,20$ |
| 6 | .7198 | .734 | .7245 | .731 |
| 14 | .5056 | .501 | .491 | .579 |
| 20 | 1.6270 | 1.6295 | 1.649 | 1.671 |
| 51 | -.5404 | -.555 | -.564 | -.591 |
| 59 | -.6478 | -.634 | -.609 | -.602 |
| 77 | .7026 | .700 | .683 | .678 |
| 80 | 1.0919 | 1.0835 | 1.086 | .992 |

## VI. COMPUTATIONS

After the main features of the programs are presented, a particularly important part of the computation is discussed in further detail. The discussion of this computation will be used to describe our present method for combining observations of examinees on different ability levels and obtaining a rough index of goodness of fit of the equations of the periodic procedure.

### VI.1 Main Features

We begin by restating the key equations of Section II in a more compact and convenient form.

Suppose

$$\text{(VI.1)} \qquad \begin{cases} f(\cdot) = u^{-1}[u(\cdot) + b_i - b_j] \quad , \\ g(\cdot) = u^{-1}[u(\cdot) + b_k - b_\ell] \quad , \end{cases}$$

for some increasing, continuous $u$ and positive constants $b_i - b_j$, $b_k - b_\ell$.

Let $w$ be any continuous, strictly increasing solution of Abel's equation

$$\text{(VI.2)} \qquad g(\cdot) = w^{-1}[w(\cdot) + 1] \quad .$$

Define $\phi$, $\Theta$ by

$$\text{(VI.3)} \qquad \phi = wfw^{-1} \quad ,$$

(VI.4) $\quad \theta(x) = \phi(x) - x$ .

Then since $\theta(x) = \theta(x+1)$ and $\phi(x) = \theta(x) + x$ , $\phi^n(x)$ can be computed recursively for all $n$ and $x$ from any unit-length segment of the graph of $\theta$ .

For each positive integer $n$ let $m_n$ be the smallest integer $m$ such that

(VI.5) $\quad \phi^n(x) \leq x + m$

for all $x$ . Then

$$B = (b_i - b_j)/(b_k - b_\ell) = \lim m_n/n \quad .$$

In fact

(VI.6) $\quad 0 \leq B - m_n/n < 1/n$ .

Instead of (VI.5) we will use the statistically more stable condition

(VI.5') $\quad \int_0^1 \phi^n(x)dx \leq \frac{1}{2} + m$ .

This concludes the summary of Section II.

Function composition is the basic operation of our programs. Most of our functions are strictly increasing, continuous, piecewise linear mappings of the unit interval onto itself. We encoded each by listing the coordinates of 101 points on its graph. The first point is always the origin (0,0), the last always (1,1).

The first step in the implementation of the periodic procedure has already been described. It is the computation of the functions $f_{ij}$ and $f_{k\ell}$ that play the role of $f$ and $g$ in (VI.1).

The second step is solving Abel's equation (VI.2). The basic algorithm was described and illustrated in detail in Levine (1970, Section III.2; also 1975). That algorithm begins with the choice of a continuous function defined on a short interval in the domain of $g$ and then uses Abel's equation to recursively extend the initial function.

We experimented with linear initial functions and a function suggested by an asymptotic result of Krantz. We found that the recursive extension of $w$ was generally smoother and easier to encode when the initial function was proportional to

$$\int_{x_0}^{x} \frac{dt}{f_{k\ell}(t) - t}$$

where $x_0$ is such that $f_{k\ell}(x_0)$ has been computed from a large sample of candidates and $x_0 \leq x \leq f_{k\ell}(x_0)$ . The rationale for this formula can be found in Krantz (1971, page 593).

We recommend choosing items $k\ell$ such that $f_{k\ell}$ has few fixed points, since the recursive procedure defines $w$ only between a pair of fixed points containing the initial segment (Levine, 1970, Section III.2). We also recommend choosing functions $f_{k\ell}$ such that it is possible to begin the recursive procedure on an interval with statistically reliable estimates of function values. We located such intervals by counting the number of independent observations directly contributing to the estimate of each function value.

The next step is to compute the functions $\phi$ and $\Theta$. To do this we first used the solution to Abel's equation and $f_{ij}$ to define $\Theta$ by $\Theta(x) = wf_{ij}w^{-1}(x) - x$. This empirical function is of course not precisely periodic. By approximating it by a periodic function $\tilde{\Theta}$ we were able to combine information from examinees at different ability levels and simultaneously measure the extent to which the data from items $ijk\ell$ agree with the model. A discussion of $\tilde{\Theta}$ is given in the next section. Basically $\tilde{\Theta}$ is a trigonometric polynomial optimizing a measure of periodicity.

We also obtained remarkably accurate results by superimposing and averaging cycles of $\Theta$ to define an average function $S$. But two problems eventually forced us to abandon $S$ and develop $\tilde{\Theta}$: (1) For some data the mapping $x \rightarrow S(x) + x$ is nonmonotonic and (2) we failed to find a satisfactory way of averaging to assure $S(0) = S(1)$.

The empirical function $\phi(x) = \Theta(x) + x$ was replaced by $\tilde{\phi}(x) = \tilde{\Theta}(x) + x$ where $\tilde{\Theta}$ is the periodic approximation of $\Theta$.

To estimate $B(ij,k\ell)$ we used $\tilde{\phi}^n$ and condition (VI.5') to estimate integers $m_n$. Our estimate $\tilde{m}_n$ was the smallest integer $m$ satisfying

$$\int_0^1 \tilde{\phi}^n(x)\, dx \leq \frac{1}{2} + m \quad .$$

The estimate of B was the minimum of $\{\tilde{m}_n/n : n = 512, 513, \ldots 532\}$ .

## VI.2 Computation of $\tilde{\theta}$

The graph of the empirical function $\theta$ commonly appears smooth and approximately periodic over much of the domain of definition of the function. However, for very large and very small values of $x$ , the graph appears jagged and amorphous. We experimented with several ways to combine the values of $\theta(x)$ to obtain a smooth, periodic function. The one presented below does not give the best agreement with LOGIST, but it has the advantage of not requiring inspection of intermediate results.

All of the results reported in this paper have been computed by first computing a "weight function" $W(x)$ , which will be seen to control the contribution of $\theta(x)$ to $\tilde{\theta}(x)$ . Note that $\theta(x)$ is simply a complicated arithmetic expression in the observed proportions $\rho_{ig}$ and $\rho_{jg}$ or in the end points $f_{ij}(0) = 0$ , $f_{ij}(1) = 1$ . $W(x)$ is zero if an end point is used in the calculation of $\theta(x)$ . Otherwise $W(x)$ is the minimum denominator of the observed proportions. Thus $\theta(x)$ is based on at least $W(x)$ independent observations.

We recommend that $W(x)$ be computed recursively. To do this each table defining an $f_{ij}$ is accompanied by a list of integers (denominators or zero). When two functions are composed, the list is updated in the obvious way.

$\Theta$ is approximated by computing the first few terms of a Fourier series. Recall that the Fourier series of a continuous real function $h$ with period equal to one is the series $H$ defined by

$$H(x) = a_0 + \sum_{m=1}^{\infty} [a_m \cos(2\pi mx) + b_m \sin(2\pi mx)]$$

where the coefficients $a_m, b_m$ have the property that for each partial sum $H_n$

$$H_n(x) = a_0 + \sum_{m=1}^{n} [a_m \cos(2\pi mx) + b_m \sin(2\pi mx)]$$

the coefficients minimize the mean squared difference between $h$ and $H_n$

$$\text{(VI.7)} \qquad \int_0^1 [h(x) - H_n(x)]^2 \, dx \quad .$$

To filter irregularities in the graph of $\Theta$ we only considered partial sums of the form $H_5$. (We chose five terms after observing that the shape of $\Theta$ could easily be described with 5 terms and that with many terms pathologies like those associated with Gibb's phenomenon began to appear.)

Instead of minimizing the squared error (VI.7) we selected the coefficients of the approximation to minimize the weighted squared error $\| \ \|$ defined by

$$\text{(VI.8)} \qquad \|\Theta - \Theta^*\| = \int [\Theta(x) - \Theta^*(x)]^2 W(x) dx \quad .$$

This permits all values of $\Theta(x)$ to influence the shape of the approximation $\Theta$ , but only to the extent that they are based on observations.

Another advantage of (VI.8) is that it greatly simplifies programming and speeds computation. For very high values of $x$ there are no appropriate candidates for defining the graph of $f_{ij}$ . With (VI.8) we are free to choose any values for the graph that are easy to code and that make the compositions well defined. If this is done, the computer is able to compute any composition of functions the programmer specifies, whether appropriate data are available or not. However, the portions of graphs not based on observations result in zero $W$ and make no contribution to $\tilde{\Theta}$ . This means that these portions of the graphs are unable to affect the estimate of $B$ .

## VI.3 Measuring Periodicity

As noted in Section IV, the $\tilde{B}(ij,k\ell)$ can be combined in various ways to obtain overall estimates of $b_i$ . To select the best available estimates of $b_i$ or $B(ij,k\ell)$ it would be helpful to be able to examine some intermediate results in the computation of $\tilde{B}$ and determine how well $\tilde{B}(ij,k\ell)$ estimates $B(ij,k\ell)$ .

We experimented with various approaches to this problem. Our most promising begins with the observation that if the empirical function $\Theta$ cannot be approximated by a periodic function, then there are no simultaneously conjugate functions close to $f_{ij}$ and $f_{k\ell}$ . Thus by measuring the periodicity of $\Theta$ we may obtain a measure of the error in the periodic estimates.

To measure the periodicity of $\Theta$ we first considered

$$\underset{\Theta^*}{\text{infinimum}} \{\|\Theta - \Theta^*\| : \Theta^* \text{ is a continuous function with period equal to one}\} ,$$

where $\| \ \|$ is the functional $F \to \int [F(x)]^2 W(x)\, dx = \|F\|$ defined in the preceding section. To eliminate effects of scale we divided by $\|\Theta\|$. We approximated $\inf \|\Theta - \Theta^*\|$ by $\|\Theta - \tilde{\Theta}\|$. This gives the index of periodicity $I(\Theta) = \|\Theta - \tilde{\Theta}\| / \|\Theta\| = I[\Theta(ij,k\ell)]$.

To check the conjecture that $I(\Theta)$ is related to the error in the periodic procedure we compared $I(\Theta)$ with the measure of error $E(ij,k\ell) = |\tilde{B} - \hat{B}| / |\tilde{B} + \hat{B}|$. A full listing of our results is given in Table 3a, 3b, 3c. For fixed $k\ell$ the correlations between $I(\cdot\cdot,k\ell)$ and $E(\cdot\cdot,k\ell)$ are high ( r = .87, .83, .90 = for $(k,\ell) = (59, 77)$, (51, 77), (77, 20) , respectively). These correlations suggest a simpler relation that is actually present since there is one very large value of E and several very small values of E that are well predicted.

References

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968. Chapters 17-20.

Krantz, D. H. Integration of just noticeable differences. Journal of Mathematical Psychology, 1971, 8, 591-599.

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. Psychometrika, 1964, 29, 115-129.

Levine, M. V. Transformations that render curves parallel. Journal of Mathematical Psychology, 1970, 7, 410-433.

Levine, M. V. Transforming curves to curves with the same shape. Journal of Mathematical Psychology, 1972, 7, 410-443.

Levine, M. V. Additive measurement with short segments of curves. Journal of Mathematical Psychology, 1975, 12, 212-224.

Levine, M. V. Nonadditive analogues of the basic mathematics of additive measurement. Submitted to Journal of Mathematical Psychology, 1976.

Lord, F. M. An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020.

Lord, F. M. Item characteristic curves as estimated without knowledge of their mathematical form--a confrontation of Birnbaum's logistic model. Psychometrika, 1970, 35, 43-50.

Lord, F. M. Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. Research Bulletin 75-33. Princeton, N.J.: Educational Testing Service, 1975.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson and Lydicke, 1960.