Winter 1984

A Study of the Robustness of the Three-Parameter Item Response Model

James B. Flynn
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_etds

Part of the Educational Psychology Commons

Recommended Citation
Flynn, James B. "A Study of the Robustness of the Three-Parameter Item Response Model" (1984). dissertation, Old Dominion University, DOI: 10.25777/hvex-6j89
https://digitalcommons.odu.edu/psychology_etds/274

This Dissertation is brought to you for free and open access by the Psychology at ODU Digital Commons. It has been accepted for inclusion in Psychology Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
A STUDY OF THE ROBUSTNESS OF THE
THREE-PARAMETER ITEM RESPONSE MODEL

By
James B. Flynn
M.S. May 1979, Old Dominion University
B.S. May 1975, Lynchburg College

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
Psychology
December 1984

Approved by:

Glynn D. Coates (Director)
Dedicated to Sensei Tesshin Hamada
Acknowledgements

My training at Old Dominion University has been extensive. I have been provided numerous and diverse educational and research experiences and I am most appreciative of the many individuals who have contributed to these experiences. However, without the guidance, support and encouragement of Dr. Malcolm Ree of the U.S. Air Force Human Resources Laboratory, I would not have been able to complete my dissertation in the area of item response theory. Dr. Ree offered his support without asking for any form of recognition. He was both my mentor and friend, and I am truly appreciative of both his contributions to my doctoral dissertation research and his friendship.

The faculty and staff in the Psychology Department at Old Dominion University have also been supportive of my educational, research and work experiences, and I am grateful to them all. I give special thanks to the members of my dissertation committee, Dr. Glynn O. Coates, Dr. Albert S. Glickman, Dr. Bruce McAfee and Dr. Ben B. Morgan, Jr., for their readings of my dissertation, to Ms. Jackie Winston for the numerous times she directed my mail and telephone calls to the organizations in which I worked while obtaining my degree and to Mrs. Alayne Sawyer for her secretarial support and friendship.

Dr. Coates, my dissertation chair, and Dr. Glickman, my mentor and friend throughout my doctoral training, deserve special thanks. Dr. Coates contributed greatly to the final
draft of my dissertation and has provided a sort of "down-home" encouragement that has kept me going these last few months. Dr. Glickman has been both a friend and in the purest sense of the word, has exemplified what is meant by mentor.

Also, I take this opportunity to express my deep appreciation to my parents, Mr. Stephen Horvath and Mrs. Barbara Horvath, and my brothers, Dr. Gregory Flynn, Mr. Mark Flynn and Airman First Class Stephen Horvath, for their support and encouragement throughout my graduate training. Oftentimes they have given, when it meant that they would go without. I could never truly express how much their support has been appreciated.

At Old Dominion University, I have had many friends who have been very supportive of my doctoral training. I extend special thanks to Dr. Pete Winne, Mr. Russ Pond, Mr. Juny Samson, and Dr. Terry Dickinson for their support and friendship.

Finally, I thank Sensei Tesshin Hamada for his patience, understanding and training. Sensei has taught countless students in karate-doh and has been teacher, friend and inspirational leader to each of his students. He has used many forms to teach us one simple attitude - to never give up! If anyone contributed more to the completion of my graduate training at Old Dominion University, it was Sensei Tesshin Hamada. I could never express in words my appreciation and love for Sensei Hamada; my life will speak of my feelings and dedication to this truly magnificent teacher.
ABSTRACT

A Study of the Robustness of the Three-Parameter Item Response Model

James B. Flynn
Old Dominion University
Director: Dr. Glynn D. Coates

Simulation techniques were employed to investigate the use of the three-parameter item response model on psychological test data which violated the model's assumptions of large sample sizes, long tests and test unidimensionality. The accuracy of the person ability and item characteristic curve parameter estimates derived by the three-parameter item response model was evaluated. Data sets and distributions of person ability and item characteristic curve parameters were generated using a computer-based algorithm, AVRAM (Ree, 1980), which employs the three-parameter logistic probability equation described by Birnbaum (1968). A computer software package, LOGIST5 (Wingersky, Burton & Lord, 1982), which utilizes the three-parameter logistic probability equation, was used to derive the parameter estimates for the person response and the item characteristic curves.

The present study based its analyses on the unedited person-item data matrix. As such, the findings are somewhat inconsistent with those reported by studies employing an edited data matrix (e.g., Ree, 1979). However, these findings, as well as the use of the unedited data matrix, are much more consistent with the types of test situations likely to occur in
industrial-organizational research, where the focus of research will be the evaluation of differences in individual and group test scores as opposed to the design and construction of tests.

The results showed that the item discrimination, $a_i$, and lower asymptote, $c_i$, parameters of the item characteristic curve were both accurately recovered when small sample sizes and short tests were used, and conditions of item bias existed. The person ability parameter, $B_v$ was also accurately recovered; $B_v$ being more accurately recovered than any of the item characteristic curve parameters. The recovery of $b_i$, the item difficulty parameter, was most affected. The average absolute differences and root-mean-square errors obtained on $b_i$ were extremely large relative to those obtained on $a_i$ and $c_i$, as well as, those reported for $b_i$ elsewhere in the literature (Ree, 1979). Furthermore, the Pearson product-moment correlations obtained on $b_i$ were low, and major differences were reported for the means of the distributions of known and estimated $b_i$ parameters. Not only were the individual parameter estimates for $b_i$ not accurately recovered, but also, changes in the means of the distributions of $b_i$ were observed.

These findings show that when an unedited person-item matrix as used to estimate person ability and item characteristic curve parameters, conditions of sample size, test length and item bias negatively affect the estimation procedures of the three-parameter model. These effects are most significant for the item difficulty parameter, $b_i$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
When the samples available for item analysis are comprised of less than 2,000 examinees, and conditions of item bias exist, the practitioner of industrial-organizational psychology should consider the following: (1) use the three-parameter model, but proceed with caution, and remove from the person-item matrix those items for which sufficient parameter estimates cannot be obtained; or (2) adopt an alternative item response model which places less restriction on the types of data available for item analysis. BICAL (Wright & Mead, 1976), a one-parameter model which employs maximum likelihood procedures, is suggested for sample sizes of 1000 examinees, and PROX (Cohen, 1976), a one-parameter model that uses algebraic procedures, is suggested when samples are comprised of 500 examinees or less.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Item Response Theory</td>
<td>3</td>
</tr>
<tr>
<td>Industrial-Organizational Research</td>
<td>19</td>
</tr>
<tr>
<td>Robustness of the Parameter Estimation</td>
<td>20</td>
</tr>
<tr>
<td>Method</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>33</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>References</td>
<td>50</td>
</tr>
<tr>
<td>TABLES 1 - 13</td>
<td>55</td>
</tr>
<tr>
<td>FIGURES 1 - 4</td>
<td>69</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>73</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>95</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Sets Evaluated in this Study</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>Means and Standard Deviations for the Distributions of Known Item Characteristic Curve $a_i$, $b_i$, and $c_i$ Parameters</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Means and Standard Deviations for the Distributions of Known Person-Ability Parameters</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>Repeated Measures Analysis of Variance on the $a_i$ Item Characteristic Curve Parameter</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>Repeated Measures Analysis of Variance on the $b_i$ Item Characteristic Curve Parameter</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>Repeated Measures Analysis of Variance on the $c_i$ Item Characteristic Curve Parameter</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>Repeated Measures Analysis of Variance on the Person Ability Parameter ($B_v$)</td>
<td>61</td>
</tr>
<tr>
<td>8</td>
<td>Means and Standard Deviations of the Distributions of Known ($a_i$, $b_i$, and $c_i$) and Estimated ($\hat{a}_i$, $\hat{b}_i$, and $\hat{c}_i$) Item Characteristic Curve Parameters</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>The Means and Standard Deviations of the Distributions of Known ($B_v$) and Estimated ($\hat{B}_v$) Person Ability Parameters</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>Average Absolute Differences and Root Mean Square Errors for Each Item Characteristic Curve Parameter for 18 Data Sets</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>Average Absolute Differences and Root Mean Square Errors for the Person Ability Score for 18 Data Sets</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>Simple Regressions of the Item Characteristic Curve Parameters Estimates ($\hat{a}_i$, $\hat{b}_i$ and $\hat{c}_i$) on the Known Parameters ($a_i$, $b_i$ and $c_i$)</td>
<td>66</td>
</tr>
<tr>
<td>13</td>
<td>Simple Regressions of the Person Ability Parameter Estimates ($\hat{B}_v$) on the Known Person Ability Scores ($B_v$)</td>
<td>68</td>
</tr>
<tr>
<td>A-1</td>
<td>The Unedited Person-Item Matrix</td>
<td>88</td>
</tr>
<tr>
<td>A-2</td>
<td>The Edited and Ordered Person-Item Matrix</td>
<td>89</td>
</tr>
<tr>
<td>A-3</td>
<td>Grouped Distribution of the 10 Different Item Scores Obtained on the 34 Persons</td>
<td>90</td>
</tr>
<tr>
<td>A-4</td>
<td>Grouped Distribution of Observed Person Scores on 14 Items</td>
<td>91</td>
</tr>
<tr>
<td>A-5</td>
<td>Proportion to Logit Conversions</td>
<td>92</td>
</tr>
<tr>
<td>A-6</td>
<td>Corrected Item Difficulties From 34 Persons</td>
<td>93</td>
</tr>
<tr>
<td>A-7</td>
<td>Final Estimates of Person Ability Scores for All Possible Scores on the 14 Item Test</td>
<td>94</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>The Item Characteristic Curve</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>A Bar Graph of the Average Absolute Differences for the $a_i$, $b_i$ and $c_i$ Item Characteristic Curve Parameters on Tests which were either 0, 20 or 40 percent biased</td>
<td>71</td>
</tr>
<tr>
<td>4</td>
<td>A Bar Graph of the Average Absolute Differences for $B_y$ on tests which were either 0, 20 or 40 percent biased</td>
<td>72</td>
</tr>
</tbody>
</table>
Introduction

Many advancements in the field of industrial-organizational psychology have occurred as a result of the technological breakthroughs made in related scientific disciplines. The field of educational measurement is one such discipline where recent developments in item analysis procedures hold much potential for the analysis of data obtained in industrial-organizational research. Advocates of item response theory assert that this approach to item analysis enables the objective measurement of mental ability. These advancements in item analysis procedures are of both theoretical and social importance to the field of industrial-organizational psychology.

Psychologists have long questioned the extent to which classic item analysis techniques provide for the objective measurement of mental ability (Thorndike, 1926; Thurstone, 1925, 1927). Gulliksen (1950) observed that:

Relatively little experimental or theoretical work has been done on the effect of group changes on item parameters. If we assume that a given item requires a certain ability, the proportion of a group answering that item correctly will increase and decrease as the ability level of the group changes.... As yet, there has been no systematic theoretical treatment of measures of item difficulty directed particularly toward determining the nature of their variation with respect to changes in group ability. Neither has the experimental work on item analysis been directed toward determining the relative invariance of item parameters with systematic changes in the ability of the group tested (Gulliksen, 1950, 392-393; cited in Wright and Stone, 1979, viii).
Gulliksen was addressing the limitations placed on psychological testing by the measurement characteristics of classic item analysis techniques (cf, Guilford, 1954).

When classic item analysis techniques are used for the purpose of item calibration and person ability measurement, the obtained estimates of individual item and person parameters are directly affected by changes in the underlying distributions of person ability and item difficulty. Estimates of item difficulty are affected by changes in sample ability, and the obtained test scores (i.e., person ability estimates) vary as a function of changes in the difficulty levels of the items administered to individual examinees. These conditions represent a lack of measurement objectivity, and this lack of objectivity has negated both the generalization of test scores beyond a single test situation and the development of a theoretical science of mental abilities (Angoff, 1960; Loevinger, 1947; Tucker, 1953). More importantly, it has greatly restricted the ways in which test scores can be used in the management and development of employees.

Wright (1967) explains that objective measurement places three primary requirements on test instruments and procedures: (1) person ability scores are estimated independent of the measurement instrument; (2) item parameter estimates are derived independent of the calibration sample's underlying ability distribution; and (3) evaluation of how well a given set of observations can be transformed into objective measurement is independent of the specific nature of the test.
items and calibration samples on which the data are obtained. Person-free item calibration and item-independent person measurement are the necessary conditions for the generalization of measurement beyond a single set of test items or calibration sample (Wright, 1977).

The mathematical formulations of item response models provide sample-free item calibration and item-independent person measurement, and these advancements in item analysis represent important advantages over classic item analysis techniques (Anderson, Kearney, & Everett, 1968; Wright & Stone, 1979). The one- and three-parameter item response models have been the most frequently studied and a considerable amount of research has been directed at both the development of these procedures and tests of their "fit" to both empirical and simulated test data (Andersen, 1972, 1973; Andersen, Kearney & Everett, 1968; Hambleton, 1969; Hambleton & Traub, 1973; Lord, 1968, 1970; Ree, 1979; Rentz & Rentz, 1978). Numerous studies have employed these models for purposes of test design and validation (cf., Hambleton & Cook, 1977; Wright, 1977a), and industrial-organizational studies have adopted this approach for the purpose of evaluating data obtained on different groups (Guion & Ironson, 1983).

Item Response Theory

Item response theory was first introduced to educational specialists nearly 30 years ago (Lord, 1952). The most frequently researched models have been the one-parameter or Rasch model (Rasch, 1980) and the three-parameter model
(Birnbaum, 1968). The one-parameter model gained its initial support because of its statistical eloquence and mathematical simplicity (Wright, 1977b). However, the three-parameter model has become the more frequently used because of its theoretical completeness, its performance in empirical studies and its general acceptance by the testing, military and government communities (Cole, 1980).

Item response theory defines an approach to the design and validation of psychological tests which is an alternative to that offered by classic item analysis procedures. The importance of these procedures to both psychological measurement and industrial-organizational research is the objectivity of the parameter estimates derived by these models. The mathematical formulations of this approach to item analysis derive independent estimates of person and item parameters. These procedures objectively transform data obtained on psychological tests onto unidimensional scales of measurement. Person ability scores are estimated without regard to the difficulty levels of the items administered to individual examinees and item parameter estimates are unaffected by changes in sample ability.

Item response models delimit what happens when persons and items at diverse ability and difficulty levels interact along a single item-response continuum. The basic tenets of this approach to item analysis specify the probability of an examinee at a given ability level correctly responding to test items at different levels of difficulty to be a normal ogive.
function of the examinee's ability level. The greater the person's ability, the greater the probability that the person will pass any single test item (Lord, 1952). Estimates of individual item parameters are derived such that the greater the item's difficulty the less the probability of any given examinee passing the item. Persons of greater ability will, on the average, have a greater probability of success on any single test item than will less able persons, and the probability that an examinee will answer an item correctly will, on the average, be greater for an easy item that it will be for a more difficult one.

The mathematical algorithms of item response models describe the pattern of data obtained on calibration samples at different underlying ability distributions and test items at different levels of difficulty. This mathematical form takes the shape of a monotonically increasing function, with the probability of a person passing any single test item being equal to the distance between the person's and the item's location on the item-response continuum.

The form of this continuum is modeled by the mathematical algorithms of item response item analysis procedures, and these algorithms will differ depending on whether a one-parameter or three-parameter model is used. If a one-parameter model is adopted, these algorithms will differ depending on whether algebraic or maximum likelihood procedures are employed. However, the basic theoretical tenets and form of the model remain the same.
Both the one- and three-parameter models derive a single person ability parameter estimate ($\hat{B}_v$) to locate the examinee on the item-response continuum. However, they differ in the number of parameters estimated to define the mathematical form of the item characteristic curve (Lord & Novick, 1968). The Rasch or one-parameter model derives a single parameter estimate to describe the item characteristic curve. This parameter estimate is $\hat{b}_i$, and it defines the item's difficulty level. The three-parameter model derives estimates of item difficulty ($\hat{b}_i$), item discrimination, ($\hat{a}_i$), and for the lower asymptote value of the item characteristic curve, ($\hat{c}_i$).

Two concepts which are common to both the one-parameter and three-parameter models are the notions of person response and item characteristic curves (cf., Lord & Novick, 1968; Wright, 1977a; Wright & Stone, 1979). The forms of the person response and item characteristic curve functions are defined by the values obtained for the person ability ($\hat{B}_v$), and item characteristic curve ($\hat{a}_i$, $\hat{b}_i$ and $\hat{c}_i$) parameter estimates.

Classic item analysis techniques describe these functions as person and item $p$-values, that is, the percentage of items correctly answered by an individual examinee and the percentage of examinees which miss or incorrectly respond to a single test item. However, item $p$-values vary as a function of changes in the underlying ability distributions of the calibration samples, and person $p$-values fluctuate depending on whether an easy or difficult test is used to estimate the person ability. Because the mathematics of classic item analysis procedures do
not correct the derived person and item p-values for effects due to differences in the underlying distributions of item difficulties and sample ability, they fail to obtain objective estimates of person ability scores and item characteristic curve parameters.

In contrast, the person response and item characteristic curve functions derived by item response procedures are not affected by changes in the underlying distributions of item difficulties and sample abilities. The mathematical algorithms of these procedures correct these functions for differences in the means and standard deviations of these underlying distributions. This is what is meant by item-free person measurement and sample-free item calibration, and these are necessary conditions for measurement objectivity (Wright, 1977b).

Figure 1 illustrates a person response curve (Wright & Stone, 1979). The person response curve is described by a single parameter, $B_v$, which defines the person's ability level and locates the individual examinee on the item-response continuum. This location is that point on the item-response continuum where the examinee correctly responds to test items 50 percent of the time.

The person response curve takes the form of a monotonically increasing function. The number of items correctly answered by an examinee at a given ability level increases as the difficulty level of the items administered to the examinee decreases. When the person's ability level, $B_v$, is less than
the item's difficulty level, $b_i$, the probability that the examinee will correctly respond to the test item will be something less than 0.50. When $B_v$ and $b_i$ are equal, that is, the person and item are located at the same point along the item-response continuum, the examinee has a fifty-fifty chance of correctly responding to the test item. When the person's ability level is greater than the difficulty level of the test item, the person will have something better than a fifty-fifty chance of correctly answering the test item. The probability that any single person will correctly respond to a given item, that is, $P(X_{vi} = 1 | B_v, b_i)$, is a direct function of the absolute difference between the locations of the person, $B_v$, and the item, $b_i$, on the item-response continuum, that is, $(B_v - b_i)$.

-- Insert Figure 1 about here --

The person response curves obtained by item response procedures are similar to the response curves derived by the psychophysical method of limits. However, rather than locating the response limen or absolute threshold at a point on a psychological continuum where the subject detects the stimulus object 50 percent of the time, item response procedures identify the examinee's ability level as that point on the item-response continuum where the examinee passes test items 50 percent of the time. Still, the basic paradigm is the same. Examinees are located on the continuum at that point where their response pattern indicates maximum confusability.
Like those response curves obtained in psychophysical research, the person response curve takes the form of a monotonically increasing function. While this algebraic function or response curve can be described by any of a number of cumulative distribution functions, the majority of item response models utilize the logistic function to transform the normal ogive function depicted in Figure 1 into a linear scale of measurement (Baker, 1977). This linear function defines the unit of measurement for person ability to be the logit. Person ability scores typically vary between -2.5 and +2.5 logits.

In addition to the person response curve, item response procedures also derive item characteristic curve functions. Figure 2 employs the three-parameter model to illustrate the item characteristic curve function. This function is described by the values obtained for the three item characteristic curve parameters: \( a_i \), \( b_i \) and \( c_i \).

---

Insert Figure 2 about here

---

The item discrimination parameter, \( a_i \), defines the slope of the item characteristic curve function. It describes the item's ability to differentiate test performance at a specific level of ability. Values for \( a_i \) range from 0.50 to 2.50. A typical value for \( a_i \) is 1.0. Values below .50 indicate insufficiently discriminating items, and values above 2.0 are found infrequently (Ree, 1979).

The item characteristic curve parameter, \( b_i \), defines the item's difficulty level. It locates the item characteristic
curve on the item-response continuum at a point where examinees miss the test item 50 percent of the time. Values typically obtained for $b_i$ range from -2.5 to +2.5 logits.

The $c_i$ parameter describes the lower asymptote of the item characteristic curve. It restricts the lower tail of the curve to a level greater than zero. This avoids the case of persons at extremely low ability levels having a zero probability of making a correct response (Baker, 1977). Values for $c_i$ range from 0.0 to .30 logits (Ree, 1979) and are typically smaller than the values which would be obtained if persons of low ability were to guess randomly on the item (Hambleton, Swainathan, Cook, Eignor & Gifford, 1978).

Unlike the three-parameter model, the one-parameter model derives a single parameter estimate to describe the item characteristic curve function depicted in Figure 2. This single parameter estimate is $\hat{b}_i$; it describes the item's difficulty level and locates the item on the item-response continuum.

The one-parameter model assumes all item discrimination parameters have a value of 1.0 logits. And while this may not be a valid assumption for paper-and-pencil test administration procedures, computer-based adaptive testing procedures do provide a testing environment in which $a_i$ can remain constant at 1.0 logits. If person ability scores are based only on those items which provide maximum information about the examinee's location on the item-response continuum, that is,
items to which examinees correctly respond 50 percent of the
time, then the item discrimination values of all item
characteristic curves would be equal to 1.0 logits.

In addition to the assumption that item discrimination
values remain constant at 1.0 logits, the one-parameter model
assumes that guessing does not occur in the test situation;
more correctly stated, it assumes that examinees will not be
administered items which are extremely difficult or easy
relative to the examinee's ability level. Person ability
scores are obtained only on those items were maximum
confusability exists. Person ability scores are based on the
examinee's responses to items whose difficulty levels are equal
to the person's ability level. When adaptive testing
procedures are employed, estimates of person ability are not
based on items whose difficulty levels are extreme relative to
the examinee's ability level, that is, extremely difficult or
easy test items. Adaptive testing procedures provide a
testing context in which guessing, as defined by item response
models, does not occur.

Although item response models are relatively simplistic in
their basic theoretical tenets, their mathematical algorithms
are extremely complex. Wright and Stone (1979) offer the most
straightforward explanation of the basic theoretical tenets and
methodological procedures of item response models. They write
that the probability of a person correctly answering an item is
determined by the distance between the person's and the item's
location on the item-response continuum, that is, the difference between the person's ability level and the item's difficulty level.

This difference \((B_Y - b_i)\), as noted by Wright and Stone (1979), varies between minus and plus infinity, indicating that calibration samples and test items can be located at opposite ends of a single item-response continuum. For the statement \(\{X_{vi} = 1 \mid B_Y, d_i\}\) to be probabilistic, (i.e., \(P \{X_{vi} = 1 \mid B_Y, b_i\}\)), it is necessary that the difference \((B_Y - b_i)\) vary between zero and one, and not minus and plus infinity. This is accomplished by applying the difference \((B_Y - b_i)\) as an exponent to the natural constant \(e\), where \(e = 2.718.28\). This yields the statement, \(e^{(B_Y - b_i)}\), which can be converted to the form \(\exp (B_Y - b_i)\). The term \(\exp (B_Y - b_i)\) varies between zero and plus infinity. Forming the ratio \(\frac{\exp (B_Y - b_i)}{1 + \exp (B_Y - b_i)}\), causes the difference \((B_Y - b_i)\) to vary between zero and one, yielding a probabilistic statement about a person at a specific level of ability passing a test item at a given level of difficulty. The probability of a correct response is specified as:
\[
\frac{e^{Dai \cdot (Bv - bi)}}{1 + e^{Dai \cdot (Bv-bi)}} \quad \text{(Equation 1)}
\]

where:

\[ P\{Xvi = 1|Bv,bi\} \] is the probability of person \( Bv \) correctly responding to item \( bi \), given ability \( v \) and item difficulty \( i \),

\( Bv \) is the ability level of person \( v \),

\( bi \) is the difficulty level of item \( i \),

\( \bar{ai} \) is the average value assigned for item discrimination and is equal to 1.0,

\( D \) is a constant equal to -1.7, and

\( e \) is the natural constant and equals 2.71828.

Equation 1 describes the one-parameter or Rasch model. The majority of one-parameter models utilize maximum likelihood procedures to simultaneously arrive at estimates of person ability, \( Bv \), and item difficulty, \( bi \). Estimates of person ability and item difficulty are based on the continued product of the one-parameter model (Equation 1) over all the persons and items contained in the person-item data matrix. Logarithmic conversions are performed on the obtained likelihood values and through a series of iterative procedures person and item parameter estimates are adjusted for the local affects of the underlying distributions of sample ability and
test difficulty. These adjustments amount to a partialling out of variance in the obtained person ability scores and item difficulty estimates which is attributable to differences in the means and standard deviations of the underlying distributions of test difficulty and sample ability. Wright and Stone (1979, pp. 62-65) provide a detailed presentation of the mathematical equations of a one-parameter model based on maximum likelihood procedures.

In addition to this presentation, Wright and Stone (1979, pp. 28-45) also describe an alternative approach to one-parameter item analysis which is based on algebraic procedures. Like those models based on maximum likelihood procedures, algebraic formulations of the Rasch model correct the estimates of person abilities for affects which are due to differences in the means and standard deviations of the distributions item difficulties. Also, they correct item difficulty estimates for variance which is attributable to differences in the means and standard deviations of the underlying distributions of sample ability.

Research which has evaluated the use of these algebraic procedures for the analysis of test data has reported that they provide accurate approximations of the item and person parameter estimates obtained with the more sophisticated maximum likelihood procedures (cf. Wright, 1977a). More importantly, these procedures place less restriction on the size of the calibration samples necessary for item analysis, with samples smaller than 500 examinees being appropriate for
item analysis. Because these algebraic procedures place less restriction on the size of the calibration samples necessary for item analysis, it is likely that these procedures (cf. Cohen, 1976; Wright & Stone, 1979) will be the ones most frequently employed in industrial-organizational studies. For this reason and because these procedures offer a straightforward approach to describing what is meant by sample-free item calibration and item-independent person measurement, the procedures and algorithms reported by Wright and Stone (1979, pp. 28-45) have been edited and reproduced in Appendix A.

The three-parameter model is the one most frequently employed by the military, government agencies, and the testing industry. It is the most mathematically complex model and it makes the greatest assumptions about the types of data appropriate for item analysis. The three-parameter model was formulated on the basis of the earlier two-parameter model of Birnbaum (cf., Lord & Novick, 1968). Like the one-parameter model, the three-parameter model derives a single person ability parameter, $B_v$, to locate the examinee on the item-response continuum. However, using the three-parameter model, the test constructor is tasked with estimating two additional item characteristic curve parameters for each of the $n$ test items. In addition to $b_i$, the three-parameter model derives estimates for an item discrimination parameter, $a_i$, as well as, estimates for the lower asymptote value of the item
characteristic curve, ̂ci. Equation 2 presents the mathematical form of the three-parameter model (Hambleton, Swaminathan, Cook, Signor & Gifford, 1978).

\[ P \{Xvi = 1 \mid Bv, ai, bi, ci\} = \frac{e^{Da1 (Bv-bi)}}{1 + e^{Da1 (Bv-bi)}} (Equation 2) \]

where:

\[ P \{Xvi = 1 \mid Bv, ai, bi, ci\} \] is the probability of person v correctly responding to item i given the values Bv, ai, bi and ci,

Bv is the ability level of person v,
ai is the item discrimination value for item i,
bi is the difficulty level of item i,
ci is the lower asymptote of item i,
D is a constant equal to -1.7, and
e is the natural constant and equals 2.71828.

For the majority of persons adopting this approach, the mathematics of the three-parameter model remain encoded in the algorithms of the individual computer software package chosen for use. Lord and Novick (1968) present a detailed study of these procedures and Wingersky, Barton and Lord (1982) describe the iterative steps which comprise the most recent of three-parameter item analysis procedures, LOGIST5.

The three-parameter model uses maximum likelihood procedures to obtain the model's parameter estimates. It bases these values on the continued product of Equation 2 over all person, v, and item, i, values included in the person-item
matrix. The inclusion of $a_i$ and $c_i$ in the three-parameter model increases both the procedures mathematical complexity and the number of steps employed to simultaneously solve the mathematical equations used to estimate person ability scores and item characteristic curve functions.

Wingersky, et al (1982) provide the most interpretable reading of these procedures. Wingersky describes these estimation procedures as being categorized into four steps with different sets of parameters being estimated in each step. In the first step, examinee ability ($B_v$) and item difficulty ($b_i$) are estimated. The item discrimination parameter ($a_i$) and lower asymptote value ($c_i$) remain fixed. In the second step, $B_v$ is fixed and all three item characteristic curve parameters are estimated. In the third step, $B_v$ and $b_i$ are again estimated and $a_i$ and $c_i$ are fixed. In the fourth and final step, person ability, $B_v$, is fixed and all three item characteristic curve parameters are estimated.

Each step is comprised of several stages, and within each stage the person and item characteristic curve parameters are estimated one at a time. To obtain sufficient estimates for the person and item characteristic curve parameters several iterations are typically required before the parameter estimates are reliably reported. A step is considered converged when the increase in the criterion (i.e., the maximum likelihood function) between two successive stages is less than the specified percentage for that step. If these procedures fail to arrive at sufficient estimates, default values are
assigned. For instance, if sufficient estimates of person ability are not obtained, a default value which would be equal to the lower or upper limit ($\pm$ 2.5 logits) of person ability scores would be assigned for person $v$ (Wingersky, et al., 1982, pp. 1, 3, 4).

Although the one- and three-parameter models do differ in terms of the mathematical algorithms which comprise these procedures, the theoretical tenets and mathematical methods of these models are the same. Both models correct item and person $p$-values for effects due to the underlying distributions of test difficulty and sample ability, and both approaches convert these estimates to logarithmic scales. Maximum likelihood procedures are used by both models to estimate the item characteristic curve parameters and person scores. However, the one-parameter model derives a single parameter estimate to describe the item characteristic curve function and the three-parameter model obtains three parameters to describe this same curve.

The derivation of $a_i$ and $c_i$ item characteristic curve parameters enables the three-parameter model to describe empirical test data better (Cole, 1980). However, obtaining these additional item parameters also makes the three-parameter model more mathematically complex. Advocates of the one-parameter model report that the item discrimination and lower asymptote value parameter estimates obtained with the three-parameter model are insufficiently derived (Andersen, Kearney, Everett, 1968; Wright, 1977; Wright & Stone, 1979).
Furthermore, Wright (1977) adds that these additional parameters should be controlled in the testing context and not included in the model of psychological test performance. In addition to these conceptual concerns, it should be noted that the inclusion of these parameters requires that large sample sizes be employed for the purposes of item calibration. This restriction is probably the greatest criticism of the three-parameter model.

**Industrial-Organizational Research**

Item response theory has begun to receive considerable attention in the field of industrial-organizational psychology. These procedures have been applied to the evaluation of item bias (Craig & Ironson, 1981; Draba, 1977; Durovic, 1975a, 1975b; Ironson, 1981; Ironson & Subkoviak, 1979; Lord, 1977b; Rudner & Convey, 1978; Rudner & Getson, 1980; Shepard & Camilli, 1980), the study of adverse impact (Ironson, Guion & Ostrander, 1982; Raju & Edwards, 1983), and the analysis of Likert-type questionnnaire data (Andrich, 1978; Hulin, Drasgow & Komocur, 1982; Parsons, 1983; Parsons & Hulin, 1982; Wright & Masters, 1980). Guion and Ironson (1983) paint a very optimistic picture for this approach, and report the following potential applications for these procedures to data obtained in industrial-organizational studies: (1) the development of diagnostic-continuous testing programs; (2) the objective evaluation of group differences in perceived job characteristics across organizations and organizational levels; (3) the study of differences in worker traits (e.g., work
motivation, job involvement, worker autonomy); (4) the evaluation of employee growth and development; and (5) the objective analysis of training effects.

Needless to say, these applications represent much potential for major advancements to be made in the methods and procedures of industrial-organizational research. More importantly, they hold much significance for the management and development of employees within the context of business and industry. However, questions remain regarding the applicability of these procedures to the types of data obtained in industrial-organizational research. Studies are needed to determine the extent to which these procedures are robust in the face of violations of model assumptions.

Robustness of the Parameter Estimation Procedures

Hambleton and Cook (1977) note that item response models make at least three fundamental assumptions about the types of data available for item analysis. The first assumption is that the test or psychological inventory measures a unidimensional construct or latent-variable continuum; estimates of individual item characteristic curve parameters must remain stable across different groups or calibration samples. This means that data obtained for a single set of test items on different calibration samples must define a single scale of measurement.

The second assumption of item response models is the local independence of individual test items. Item response models assume that an examinee's response to a single test item is not
affected by the responses made to other test items. Hambleton and Cook (1977) note that this assumption is actually an alternative form of the model's assumption of unidimensionality.

The final set of assumptions are specified by the model's mathematical form and are different for the one- and three-parameter models. These assumptions refer to the numbers of examinees and test items necessary for item calibration and person measurement. Both the one-parameter (Rasch, 1980) and three-parameter (Lord, 1952) models require relatively large calibration samples and long item vectors to arrive at sufficient estimates of the models' person and item characteristic curve parameters.

Several studies have evaluated the effects of sample size and test length on the estimation procedures of the one- and three-parameter models (Hambleton & Cook, 1980; Halín, Lissak & Drasgow, 1981; Ree, 1981; Ree & Jensen, 1980a, 1980b; Traub, 1983). Studies which have addressed the effects of biased items on the estimation procedures of item response models (Angoff & Ford, 1973; Goldstein, 1980) generally conclude that these procedures are unaffected by the inclusion of biased items on the person-item matrix. Unless the extent of item bias results in a test comprised of several factors with equally large factor loadings, item bias is typically not a problem. Factor analytic techniques are suggested to determine the extent of item bias and if item bias is a problem, multidimensional models are available to handle these types of data (cf. Traub & Lam, 1985; Wright, 1977a). Sample sizes of
500 examinees have been reported sufficient for the one-parameter model, but sample sizes of at least 2,000 examinees have been typically reported to be necessary for the three-parameter model. Also reported, tests comprised of 60 items have been shown generally acceptable for the one-parameter model, and tests comprised of 80 items have been suggested for the three-parameter model. (Hambleton & Cook, 1980; Hulin, Lissak & Drasgow, 1981; Lord, 1979b; Ree & Jensen, 1980a, 1980b). However, no studies have systematically evaluated how the three-parameter model is affected when sample sizes are less than 2,000 examinees or tests are comprised of fewer than 80 items. If the estimation procedures of the three-parameter model are not robust to the assumptions of large sample sizes and long tests, it is likely that these procedures will not be applicable to data obtained in industrial-organizational studies.

To date, the three-parameter model has been the approach most frequently employed by industrial-organizational research. However, before it is wholeheartedly adopted by industrial-organizational psychologists, the effects of violating the model's assumptions of test unidimensionality, large sample sizes and long item vectors need to be systematically evaluated. The present study evaluates the applicability of the three-parameter item response model to data which is more typical of industrial-organizational research. It provides a test of the robustness of the model's person ability and item characteristic curve parameter.
estimates to violations of the model's assumptions of unidimensionality, sample size and test length. The three-parameter model (Lord, 1952) was chosen for investigation because of its general acceptance by both the military and testing communities, and because it is the most frequently researched model.

The present study evaluates three specific null hypotheses about the effects of violating the assumptions of the three-parameter model:

Hypothesis 1: Sample size does not affect the recovery of known person and item parameters,

Hypothesis 2: Test length does not affect the recovery of known person and item parameters, and

Hypothesis 3: Item bias, (i.e., violations of the model's assumption of unidimensionality) does not affect the recovery of known person and item parameters.
Method

The present study investigated the effects of violations of the three-parameter model's assumptions of test unidimensionality, sample size and test length on the recovery of known person ability scores and item characteristic curve parameters. The known person ability scores and item characteristic curve parameters, as well as, person-item data matrices, were generated with simulation techniques which employed the three-parameter logistic probability equation described by Birnbaum (1968). These computer-based algorithms, AVRAM, have been described elsewhere by Ree (1980).

Person response curves were obtained on the two distributions of known person ability scores which were simulated in this study (i.e., Group 1 and Group 2 examinees). In generating these distributions, the mean ability of Group 1 examinees (μv = 0.0 logits, SD = 1.0 Logits) was set at approximately one standard deviation greater than that of Group 2 examinees (μv = -0.8 logits, SD = 0.8 logits). Additionally, the distributions of Group 1 person ability scores were simulated to be more dispersed than the distributions Group 2 ability scores. The procedures used to simulate the unbiased item characteristic curve parameters set the ai parameter of the item characteristic curve function at 1.000 logits (SD = 0.300 logits), the bi parameter at 0.000 logits (SD = 1.000 logits) and the ci parameter at 0.200 logits (SD = 0.050 logits).
Violation of the three-parameter model's assumption of test unidimensionality was simulated by biasing the item discrimination, $a_i$, and item difficulty, $b_i$, parameters of the known item characteristic curve functions. The biased test items were obtained from the known item characteristic curve functions based on 1000 cases of Group 1 examinees and test lengths of 40, 80, and 1000 items. One thousand cases and not 500 cases were employed, because it was hypothesized that the larger sample sizes would provide more stable parameter estimates. The determination of which item characteristic curves were to be biased was based on the inspection of the reliability of the item characteristic curve parameter estimates obtained on the unbiased data sets. If LOGIST5 failed to arrive at sufficient estimates (i.e., default values for parameter estimates had been assigned) the item was not biased. Item bias effects would not be interpretable for these items.

Biased item characteristic curves were uniformly distributed at all levels of item difficulty. The biased item characteristic curve functions were developed by manually editing the $a_i$ and $b_i$ parameters of the known item characteristic curves. The $a_i$ and $b_i$ parameters of the biased item characteristic curves were edited so that these values for biased items would be greater than twice the standard error of the parameter estimates obtained on the unbiased data sets. This method of simulating item bias was thought reasonable because: (1) it offered a straightforward approach to the
simulation of item bias (i.e., bias was simulated by affecting the item difficulty and item discrimination values of the item characteristic curve functions); (2) it was consistent with item response methods of detecting bias, in which item bias is identified by evaluating the residual variances of item characteristic curve parameter estimates obtained on different groups; and (3) it was reasonable to assume that since the additive transformations of the biased test items were equal to twice the standard error of the estimates obtained on the unbiased data sets, the simulated item bias could not be attributed to chance alone.

Six sets of unbiased data were simulated. Twelve data sets were developed which included different numbers of biased test items. The data which comprised the unbiased person-item matrices were obtained on the simulated cases of Group 1 examinees. The data which comprised the biased matrices included data obtained on the simulated cases of Group 1 examinees that were administered unbiased tests of different lengths and the simulated cases of Group 2 examinees that received tests that contained different numbers of biased items.

The three components of the study's design were varied as follows: sample size was 500 or 1000 cases; test length was 40, 80 or 100 items, and violations of the model's assumption of test unidimensionality was simulated by biasing for Group 2 examinees the item discrimination, \( a_i \), and item difficulty, \( b_i \), parameters of either 20 or 40 percent of the items included in biased tests.
Table 1 describes the 18 data sets evaluated in this study. These data sets were generated by the different combinations of the specified levels of sample size, test length and the number of biased items which comprised a single item vector. Differences in the underlying ability distributions of the simulated cases of Group 1 and Group 2 examinees were developed so that the data sets would more realistically resemble data obtained in empirical studies. However, these differences were not evaluated in this study.

Insert Table 1 about here

Table 2 presents the means and standard deviations of the distributions of known item characteristic curve parameters. These parameters were simulated with AVRAM (Ree, 1980).

Insert Table 2 about here

Table 3 presents the means and standard deviations of the distributions of known person ability scores simulated in this study. Again, AVRAM (Ree, 1980) was employed for purposes of data simulation.

Insert Table 3 about here

A three-parameter model, LOGIST5 (Wingersky, Barton & Lord, 1982) was used to estimate the person ability and item characteristic curve parameters for each of the 18 simulated
person-item matrices. Estimation of the known person ability scores and item characteristic curve parameters were based on the unedited person-item data matrix. Although this method was not consistent with the approach typically adopted by studies using item response, item analysis procedures in the design and construction of tests, it was thought reasonable because the focus of this study was the feasibility of applying these procedures to data obtained in industrial-organizational studies. In this context, editing of the person-item data matrix is not likely to occur because the focus of research is the evaluation of individual and group differences in examinees' responses to individual test items and not test construction, and the tests or inventories typically used are standardized, commercially available instruments and removing items from the data matrix is not often considered.

The person ability and item characteristic curve parameter estimates obtained with LOGIST5 were then compared with the known parameters to evaluate the robustness of the three-parameter model to violations of model assumptions. The effects of sample size, test length, and degree of item bias were evaluated for the recovery of both the known person ability scores (Bv) and item characteristic curve parameters (ai, bi, and ci). Item characteristic curve parameter estimates (âi, Ûi, Ûi) were compared with known item parameters (ai, bi, and ci), and estimates of person ability (ûv) which were based on âi, Ûi, and Ûi, were compared with the known person ability scores (Bv), which had been generated by AVRAM along with the known item characteristic curve parameters and item-person matrices.
The data sets simulated in the present study were subjected to five basic methods of statistical analysis. These methods included: (1) tests of the overall effects of sample size, test length and item bias on the recovery of the individual $a_i$, $b_i$, and $c_i$ item characteristic curve parameters and the known person ability scores ($B_v$); (2) comparison of the means and standard deviations of the distributions of known and estimated item and person parameters; (3) computation of the average absolute differences (AAD) between the known and estimated parameters; (4) computation of the root-mean-square errors (RMSE) between the parameters and parameter estimates; and (5) the derivation of Pearson product-moment correlations on the parameters and parameter estimates. The rationale for computing each of these indices is described in greater detail elsewhere (Vale, Maurelli, Gialluca, Weiss & Ree, 1981) and is discussed only briefly here.

Repeated measures analyses of variance were performed to evaluate the overall effects of violations of the model's assumptions of sample size, test length and test unidimensionality, on the estimation of the individual item characteristic curve and person ability parameters.

The means and standard deviations of the distributions of known and estimated parameters were evaluated to provide an indication of the fidelity of the derived estimates. However, this approach is merely an index of the relative locations of the known and estimated parameter distributions; severe errors of estimation are often not detected. By computing the average
absolute differences (AAD) between the known parameters and the obtained estimates, errors of estimation are more likely to be detected. Although the root-mean-square error (RMSE) is similar to the average absolute difference, it more heavily weights severe errors of estimation. The root-mean-square error was computed by taking the square root of the mean of the squared differences between the true and estimated parameters. The final index, the Pearson product-moment correlation, focuses on differences in the relative positions of corresponding known and estimated parameters. It is sensitive to both the overall variance introduced into the parameter estimation procedures, as well as, extreme deviations of parameter estimates from known parameters.

AVRAM

Data were simulated by substituting the vectors of person ability scores obtained on the simulated cases of Group 1 and Group 2 examinees and the generated vectors of biased and unbiased item characteristic curve parameters into the three-parameter logistic probability equation,

\[ P(B_v) = C_i + (1 - C_i) \frac{e^{D_{ai} (b_v - b_i)}}{1 + e^{D_{ai} (b_v - b_i)}} \]  (Equation 3)

where \( P(B_v) \) is the probability of person \( v \) answering item \( i \) correctly, and \( a_i, b_i, \) and \( c_i \) are the item characteristic curve parameters.
parameters for item $i$. Equation 3 is analogous to Equation 2, but represents a shortened version of the three-parameter logistic probability equation. The generated data sets represent the probability of 1 through $v$ examinees "passing" 1 through $i$ test items, that is, $P_{iv} (0 < P_{iv} < 1)$. They form $N \times n$ matrices whose $(i, v)$th elements are $P_{iv}$.

The formulation of the item-response data matrix was based on the development of item-response vectors, where these vectors represent dichotomous test data with an incorrect response equal to 0 and a correct response equal to 1, which were then compared with the numbers, $X_{v}$, drawn from a uniform (rectangular) distribution ranging from 0.0 to 1.0. If $X_{v}$ was larger than $P(B_{v})$, an incorrect response was specified; otherwise, a correct response was specified. The obtained item-person data matrices formulated the basis for evaluating the item analysis procedures of LOGIST5. The simulated data sets provided known parameters for the generated item characteristic curves, true scores for examinee ability and specified levels of item bias. The mathematical procedures which comprise AVRAM, as well as the use of AVRAM in other research studies are documented by Ree (1979, 1980).

Data Analysis

LOGIST5 (Wingersky, et al., 1982) was used to estimate person ability scores and item characteristic curve $a_{i}$, $b_{i}$, and $c_{i}$ parameters for each of the 18 data sets. Person ability scores were estimated using maximum likelihood procedures. Maximum likelihood estimates (MLE) of $B_{v}$ are computed using the likelihood function defined in Equation 4,
where \( Q(Bv) = 1 - P(Bv) \), and \( u \) is 1 if the item is answered correctly and 0 if answered otherwise. Maximum likelihood estimates of \( Bv \) were derived from the \( \hat{a}_i's, \hat{b}_i's, \) and \( \hat{c}_i's \) (i.e., the item characteristic curve parameter estimates), as would be done in actual test administration with precalibrated items.

Similar sets of statistical analyses were performed on the known and estimated person and item characteristic curve parameters. Repeated measures analyses of variance were performed which aggregated across all data sets. The known parameters and the parameter estimates represented the two levels of the trial factor. Descriptive statistics were obtained on the distributions of known and estimated person and item parameters. Measures of absolute difference and root-mean-square error were obtained. Pearson product-moment correlations were obtained and simple regressions of parameter estimates on the known parameters were performed.
Results

Tables 4, 5, 6 and 7 report the results of the repeated measures analyses of variance (ANOVAs) on the known and estimated person and item characteristic curve parameters. These analyses aggregate across the 18 sets of item parameters reported in Table 1 for the purpose of summarizing the overall effects of sample size, test length and item bias on the estimation procedures of the three-parameter model.

The main effects of sample size (G), test length (H) and item bias (I) show the extent to which the estimation of item characteristic curve parameters on both unbiased and biased data sets is affected by violations of the model's assumptions. The "trial" factor (R) indicates the extent of the recovery of known person and item parameters by the three-parameter procedures, that is, the extent to which these procedures resulted in differences in known and estimated parameters. This basic experimental design was employed for the analysis of both the individual item characteristic curve parameters and the person ability scores.

Insert Tables 4, 5 and 6 about here

Although expected, sample size did not affect the estimation of the $a_i$, $b_i$, and $c_i$ item characteristic curve parameters. However, it did interact with the recovery of the lower asymptote value of the item characteristic curve, $c_i$, ($F = 5.852; df = 1; P < .05$) does not account for an
appreciable amount of the total variance in the estimation of $c_i$. The effects of test length were significant only for the lower asymptote value ($F = 4.231; df = 2; < .05$). Item bias affected the estimation of the item discrimination parameter, $a_i$, ($F = 3.358; df = 2; P < .05$) and the lower asymptote value, $c_i$, ($F = 3.050; df = 2; P < .05$). No bias effects were observed for the item difficulty parameter, $b_i$. However, item bias did interact significantly with the trial factor (i.e., the recovery of the known parameters) for each of the individual item characteristic curve parameters. Given that the $a_i$ and $b_i$ parameters were experimentally manipulated to be biased against Group 2 examinees, the effects for the item discrimination ($F = 31.696; df = 2; P < .05$) and item difficulty ($F = 30.585; df = 2; P < .05$) parameters were expected. The effects on the recovery of the lower asymptote value ($F = 4.906; df P = 2; < .05$) illustrate the effects which biasing $a_i$ and $b_i$ have on the recovery of an unbiased $c_i$ parameter.

Summarizing the results reported in Tables 4, 5 and 6, item bias affected the recovery of all item characteristic curve parameters (i.e., $a_i$, $b_i$ and $c_i$). The lower asymptote value, $c_i$, was most affected by the conditions evaluated in this study, with the main effects of test length and item bias, and the interaction of sample size and item bias with the trial factor, all being significant. None of the effects found to be significant accounted for appreciable amounts of the total variance in the estimation procedures of the three-parameter model, and between subjects the error terms were large relative
to those reported for the experimental conditions evaluated in the present study. It is hypothesized that this was due to the small sample sizes (less than 2,000) used in this study. When less than 2,000 examinees are used for item calibration the maximum likelihood procedures of the three-parameter model are unable to stabilize on the item characteristic curve parameters. Default values are then assigned, and these values represent extreme parameter estimates, typically deviating greatly from the known parameters. As such, the inclusion of these items in an unedited person-item matrix negatively affects the estimation procedure of the three-parameter model.

Table 7 reports the results of a repeated measures analysis of variance on the known person ability scores and parameter estimates. This analysis aggregated the 18 data sets reported in Table 1. As shown in Table 7, the conditions evaluated in this study (sample size, test length and item bias) all had a marked affect on the estimation of person ability scores. The effects of sample size ($F = 8.771; \text{df} = 1; P < .05$), test length ($F = 5.217; \text{df} = 2; P < .05$) and item bias ($F = 4.101; \text{df} = 2; P < .05$), as well as, the interaction effects, were all statistically significant. The recovery of known person ability scores was affected ($F = -49.104; \text{df} = 1; < .05$), and the interaction effects of the trial factor with sample size ($F = 2.010; \text{df} = 1; P < .05$) and item bias ($F = 15.820; \text{df} = 2; P < .05$) were also significant. Estimation of the person ability parameter was affected to a greater extent by the conditions evaluated in this study, than were any of the individual item characteristic curve parameters.
Table 8 shows the means and standard deviations of the distributions of the known item characteristic curve parameters and the means and standard deviations for the distributions of estimated item characteristic curve parameters (\( \hat{a}_i \), \( \hat{b}_i \) and \( \hat{c}_i \)). No real differences were observed for means and standard deviations of the distributions of known and estimated \( a_i \) and \( c_i \) parameters. However, for \( b_i \), the item difficulty parameter, sizeable differences, as large as 0.5 logits, existed between the means of the distributions of known and estimated \( b_i \) parameters. These differences existed for even the most robust of data sets, that is, DAT6. DAT6 was obtained on samples of 1000 examinees and 100 unbiased test items. As such, the conditions evaluated in the present study and a marked affect on the locations of the distributions of known and estimated \( b_i \) parameters on the item-response continuum.

Table 9 reports the means and standard deviations of the distributions of known and estimated person ability parameters; the known parameters were reproduced from Table 3 for purposes of comparison with the parameter estimates. Although the conditions of sample size, test length and item bias did affect the estimation of person ability, these effects were minimal.
and not readily interpretable. Looking at the unbiased data sets, the smallest mean values obtained for $B_v$ and $\hat{B}_v$ were 0.033 and 0.39 logits, respectively. The highest values obtained on the unbiased data sets were 0.049 and 0.042 logits. For the biased data sets, these differences were more extreme because the simulated cases of low ability (Group 2) examinees were included in the calibration sample. The lowest values for $B_v$ and $\hat{B}_v$ were -0.010 and -0.038 logits, respectively; the highest values were 0.040 and 0.046 logits, respectively.

One striking observation is that the means and standard deviations of the distributions of person ability estimates did not vary, except for the conditions of sample size. Data sets 1, 2 and 3 were obtained on 500 examinees and data sets 4, 5 and 6 were obtained on 1000 examinees. For the biased data sets, the first six data sets (DAT7 to DAT12) were obtained on samples of 500 examinees; the last six data sets (DAT13 to DAT18) were obtained on samples of 1000 examinees. The derived estimates remained constant across data obtained on equal sample sizes of biased and unbiased data. No explanation is offered for this observation.

Table 10 reports the average absolute differences (AAD) and root-mean-square errors (RMSE) for each of the individual item characteristic curve parameters. The AADs and RMSEs observed for $a_i$ and $c_i$ are consistent with those reported elsewhere in the literature for similar sample sizes (Ree, 1979). These values appear to be affected by the conditions of item bias,
and these effects were most readily interpretable for the average absolute differences. Figure 3 presents the average absolute differences obtained on the individual $a_i$, $b_i$ and $c_i$ item characteristic curve parameters for tests which were unbiased and for tests in which 20 or 40 percent of the items were biased. The observed AADs for $a_i$ and $b_i$ systematically increased as the number of biased items included in the test became greater; those for $b_i$ did not.

---

Insert Table 10 about here

---

Relative to the AADs and RMSEs obtained for $a_i$ and $c_i$ and to those reported for $b_i$ elsewhere in the literature (Ree, 1979), where similar sample sizes were employed, the values reported for $b_i$ in Table 10 are extremely large. The RMSEs obtained for $b_i$, suggest that these extreme differences are due to the inclusion of items in the evaluated data sets for which LOGIST5 did not arrive at sufficient parameter estimates. RMSE is sensitive to estimates which deviate severely from known parameters. Those reported in Table 10 indicate the presence of extremely poor $b_i$ parameter estimates in the data analyzed in the present study. The average absolute differences obtained for $b_i$ on tests which were unbiased and for which 20 or 40 percent of the items were biased are also presented in Figure 3. Although there were no systematic changes in the AADs for $b_i$ across the different conditions of bias, all were much greater than those reported for $a_i$ and $c_i$. 

38
Table 11 reports the average absolute differences and root-mean-square errors obtained on the person ability parameter for each data set. The findings for AAD and RMSE are similar. For the unbiased data sets, the AADs and RMSEs obtained on the person ability scores, systematically decreased, as the number of test items used for the estimation of person ability scores was increased. Similar effects were observed for the biased data sets, but within any single condition of test length, increasing the number of biased items included in the test from 20 to 40 percent tended to increase the observed AADs and RMSEs reported for Bv. The absence of extreme root-mean-square-error values suggests that relatively good estimates of person ability were obtained on all 18 data sets.

Figure 4 presents the average absolute differences obtained for Bv on tests which were 0, 20 or 40 percent biased against Group 2 examinees. The data sets graphed for the conditions of 20 and 40 percent item bias combine the person abilities of Groups 1 and 2 examinees.

The AADs obtained on Bv for data comprised of 0, 20 or 40 percent biased test items did not vary in any interpretable manner.
Table 12 reports the findings of simple regressions of item characteristic curve parameter estimates on the known parameters for each of the 18 data sets. Comparing the Pearson product-moment correlations of the $a_i$, $b_i$ and $c_i$ item characteristic parameters a systematic increase in the magnitude of the obtained correlations between the known and estimated $a_i$, $b_i$ and $c_i$ item parameters is observed as sample size is increased. Samples of 1000 examinees produced stronger correlations than did samples of 500 examinees for all item parameters. The item discrimination parameter, $a_i$, was the most accurately estimated with the obtained product-moment correlations typically being greater than $r = .60$. The estimation of $b_i$ and $c_i$ was less precise with estimated correlation coefficients of $r = .30$ or less being typical on these parameters.

Table 13 reports simple regressions of the person ability parameter estimates on the known person ability scores. Unlike the product-moment correlations obtained on the item characteristic curve parameters, those obtained for the person ability parameters were extremely high; all were above $r = .95$. Although these correlations do appear to vary
systematically as a function of test length, the changes are not significant. No effects were observed for sample size or for the number of biased items included in the test.

Insert Table 13 about here

-----------------------------
Discussion

A substantial amount of research has been performed to determine the characteristics of those test situations in which item response procedures can be employed. These studies have primarily focused on the conditions of sample size, test length and the presence of item bias (Angoff & Ford, 1973; Goldstein, 1980; Hambleton & Cook, 1980; Ree, 1981; Traub, 1983).

Sample sizes as small as 1000 have been shown to provide good item characteristic curve parameter estimates with the three parameter model (Ree, 1981) when the person-item data matrix is edited to remove those items for which sufficient parameter estimates are not obtained. Generally, this editing requires several computer runs and can result in the removal of a substantial number of items from the person-item matrix. With regard to conditions of test length, the procedures adopted by the three-parameter model are generally less affected by short tests. However, these procedures do become progressively less accurate as the number of items is reduced below 60. And finally, research has shown that unless test bias is so extreme that the item pool actually represents a multidimensional construct, the inclusion of biased items in the person-item matrix has little effect on the item analysis procedures (Angoff & Ford, 1973; Goldstein, 1980). Even so, Wright (1977) suggests the use of factor analytic procedures and the analysis of item "misfit" indices to assess the test's dimensionality; biased items can be removed from subsequent
analyses or multidimensional item response models can be employed if the bias is so great that multiple dimensions are being measured.

The purpose of the present study was to evaluate the feasibility of applying the three-parameter model to the kinds of data typically obtained in industrial-organizational studies. It provided a comprehensive evaluation of the effects of sample size, test length and item bias on the item analysis procedures of the three-parameter model, and its design provided for the systematic evaluation of both the main and interaction effects of these conditions.

Five different sets of analyses were used to evaluate the extent to which the recovery of the known person and item parameters was affected by the conditions of sample size, test length and item bias. These included: repeated measures analyses of variance, evaluation of the means and standard deviations of the distributions of known parameters and parameter estimates, the calculation of average absolute differences, the calculation of root-mean-square errors, and the study of the Pearson product-moment correlations obtained on the parameters and parameter estimates. These different analyses focused on different aspects concerning the recovery of the person and item parameters and as such, provided a comprehensive evaluation of these effects.

The repeated measures analyses of variance evaluated the overall effects of the conditions of sample size, test length and item bias. Summarizing the findings reported in Tables 4,
5 and 6, $c_i$ was most affected by the test conditions evaluated in this study. Sample size negatively affected the recovery of the lower asymptote value, $c_i$, and the estimation of $c_i$ was also affected by the conditions of test length. Item bias affected estimation of the $a_i$ and $c_i$ item characteristic curve parameters and also the recovery of $b_i$.

The conditions of sample size, test length and item bias had a noticeably greater affect on the estimation of known person ability scores, $B_y$, than they did on the estimation of the $a_i$, $b_i$ and $c_i$ item characteristic curve parameters (see Table 7). However, the significance of these effects is more likely due to the large sample sizes on which the analyses were based than it is the result of any real effects. None of these effects accounted for appreciable amounts of the variation in the estimation of $B_y$.

The conditions evaluated had little or no effect on the distributions of known and estimated person, $B_y$, and $a_i$ and $c_i$ item parameters. However, major differences were observed for the recovery of $b_i$, the item difficulty parameter.

Table 8 shows that major differences, 0.5 logits, existed between the means of the distributions of known and estimated $b_i$ parameters. These differences were observed for DAT6, the most robust of the data sets evaluated in this study. In addition to the effects on the distributions of known and estimated $b_i$ parameters, Table 10 shows that the recovery of $b_i$ for individual test items was also affected.

The root-mean-square errors reported for $b_i$, were extremely large relative to those reported for $a_i$ and $c_i$, and, these
values were twice the size of those reported for \( b_i \) in studies using similar sample sizes (Ree, 1979, 1981).

The output obtained from LOGIST5 on the \( a_i \), \( b_i \) and \( c_i \) item characteristic curve parameter estimates is presented in Appendix B. An examination of these data revealed that the procedures failed to arrive at sufficient item difficulty parameters for some test items, and it is suggested that the inclusion of these extreme values in the unedited person-item data matrix resulted in the large root-mean-square error values reported for \( b_j \) in Table 10, as well as, the large AADs reported for \( b_j \) and the differences in the means of the distributions of known and estimated \( b_j \) parameters reported in Table 8.

The results of the regression analyses and Pearson product-moment correlations reported in Tables 12 and 13 provide additional data supporting the effects of the experimental conditions of sample size, test length and item bias on \( b_i \). The item discrimination, \( a_i \), and person ability, \( B_v \), parameters were the most accurately recovered. However, \( b_i \) was again the least accurately recovered; the product-moment correlations reported for \( b_i \) in Table 12 were extremely low.

The most striking finding of the present study, therefore, was the observed effects which the experimental conditions had on the estimation of \( b_i \), the item difficulty parameter of the item characteristic curve. The means of the distributions of known and estimated \( b_i \) parameters were markedly different for each of the 18 data sets evaluated in this study (see Table
Similarly, the AADs and RMSEs reported on \( b_i \) in Table 10 were extremely large, and the Pearson product-moment correlations reported in Table 12 were extremely low. The conditions evaluated in this study affected both the locations of the distributions of \( b_i \) parameter estimates along the item-response continuum and the recovery of individual \( b_i \) parameters.

It is hypothesized that the effects observed for \( b_i \) occurred because the estimation procedures were based on the unedited person-item data matrix. Although studies have reported that the three-parameter model works reasonably well on samples as small as 1000 examinees (e.g., Ree, 1979), it needs to be recognized that this occurs only after the person-item data matrix has been edited for items on which sufficient parameter estimates can not be obtained. If an unedited matrix is used, as is more likely to be the case in the more applied industrial-organizational studies in which test construction is not typically the focus of research, researchers should expect the estimation procedures for \( b_i \) to be markedly affected by violations of the model's assumptions of sample size. These effects can be expected whenever samples comprised of fewer than 2,000 examinees are used and an unedited person-item data matrix is employed. The effects result from the inclusion of items in the person-item matrix for which the three-parameter model assigned default values as parameter estimates. The default values typically differ greatly from the known parameters, and their inclusion has a negative effect on the estimation procedures of the three-parameter item response model.
The present findings underline the need for researchers, employing the three-parameter model, to base the item analysis procedures of this approach on the edited person-item data matrix. When an unedited matrix is employed, \( b_i \) is less accurately recovered and the latent-variable continuum is unreliably defined. This is an important finding regarding the application of these procedures to the analysis of data obtained in industrial-organizational research.

Item response theory and procedures are likely to become a much more integral component of this area of research as practitioners of industrial-organizational psychology begin to understand the advantages the approach offers beyond classic item analysis techniques, and begin to become better acquainted with the theoretical tenets, assumptions and methods of item response item analysis procedures. Only recently, has research begun to address the application of these procedures to industrial-organizational psychology (cf., Guion & Ironson, 1983), and although the findings of the present study suggest that restrictions exist with regard to the application of the three-parameter model to data obtained in industrial-organizational studies, more often than not, the less restrictive Rasch model will be appropriate for these measurement contexts.

But even more importantly, item response, item analysis procedures are likely to become adopted more frequently because of their methodological, theoretical and social importance to industrial-organizational psychology, as well as, the science
of psychological measurement in general. As a method, these procedures provide objective item calibration and person measurement. They affect the measurement characteristics of the psychological scales which represent mental ability test performance.

In theory, item response models enable the design and validation of mental ability tests within the theoretical frameworks of the more process-oriented models of cognitive developmental and human cognition. Because person ability scores can be estimated independent of changes in item difficulty, different item-subsets can be administered to the same or different groups of examinees and these different calibration samples can be located on the same item-response continuum. Guion and Ironson (1983) have already discussed the importance of this measurement objectivity to industrial-organizational psychology, noting that it enables the development of diagnostic-continuous testing programs, the objective evaluation of group differences, the study of individual traits, the evaluation of individual growth and development, and the objective study of training effects. It should be noted that these applications are appropriate for both industrial and academic testing environments, and in general, item response procedures hold much potential for affecting both the testing industry and the science of psychological measurement in general.

The social implications of this approach to the item analysis of psychological test data are probably the most significant. These procedures enable the design of tests which
are diagnostic of individual skill and knowledge deficits and prescriptive of remedial training needs. They do not restrict the use and interpretation of test data to purposes of screening and classification. Data obtained from mental ability tests can be studied within the theoretical frameworks of models of cognitive development and interpreted in a manner which is consistent with notions regarding individual growth and development (cf., Glaser, 1981). Because of the methodological, theoretical and social importance of item response theory and procedures, this approach to the analysis of psychological test data will most certainly be increasingly employed by both scientists and practitioners of industrial-organizational psychology.
References


Wright, B. D., & Masters, G. N. (1980). The measurement of knowledge and attitude. Research Memorandum No. 30, Statistical Laboratory, Department of Education, University of Chicago, Chicago, IL.

Table 1

Data Sets Evaluated in this Study

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>N of Examinees</th>
<th>n of Items</th>
<th>% of Biased Items</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAT1</td>
<td>500</td>
<td>40</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DAT2</td>
<td>500</td>
<td>80</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DAT3</td>
<td>500</td>
<td>100</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DAT4</td>
<td>1000</td>
<td>40</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DAT5</td>
<td>1000</td>
<td>80</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DAT6</td>
<td>1000</td>
<td>100</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Biased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500</td>
<td>40</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT8</td>
<td>500</td>
<td>40</td>
<td>40</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT9</td>
<td>500</td>
<td>80</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT10</td>
<td>500</td>
<td>80</td>
<td>40</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT11</td>
<td>500</td>
<td>100</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT12</td>
<td>500</td>
<td>100</td>
<td>40</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT13</td>
<td>1000</td>
<td>40</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT14</td>
<td>1000</td>
<td>40</td>
<td>40</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT15</td>
<td>1000</td>
<td>80</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT16</td>
<td>1000</td>
<td>80</td>
<td>40</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT17</td>
<td>1000</td>
<td>100</td>
<td>20</td>
<td>1 and 2</td>
</tr>
<tr>
<td>DAT18</td>
<td>1000</td>
<td>100</td>
<td>40</td>
<td>1 and 2</td>
</tr>
</tbody>
</table>
Table 2

Means and Standard Deviations for the Distributions of Known Item Characteristic Curve

<table>
<thead>
<tr>
<th>DATA SETS DESIGN</th>
<th>ai Mean</th>
<th>ai SD</th>
<th>bi Mean</th>
<th>bi SD</th>
<th>ci Mean</th>
<th>ci SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>1.221</td>
<td>0.262</td>
<td>0.481</td>
<td>0.902</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>1.209</td>
<td>0.244</td>
<td>0.480</td>
<td>0.875</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>1.194</td>
<td>0.237</td>
<td>0.455</td>
<td>0.879</td>
<td>0.192</td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/40</td>
<td>1.221</td>
<td>0.262</td>
<td>0.481</td>
<td>0.902</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>1.209</td>
<td>0.244</td>
<td>0.480</td>
<td>0.875</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>1.194</td>
<td>0.237</td>
<td>0.455</td>
<td>0.879</td>
<td>0.192</td>
</tr>
<tr>
<td>Biased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>1.286</td>
<td>0.305</td>
<td>0.456</td>
<td>0.913</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT8</td>
<td>500/40/40</td>
<td>1.349</td>
<td>0.311</td>
<td>0.471</td>
<td>0.913</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/20</td>
<td>1.271</td>
<td>0.294</td>
<td>0.513</td>
<td>0.853</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>1.343</td>
<td>0.313</td>
<td>0.516</td>
<td>0.844</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/20</td>
<td>1.271</td>
<td>0.296</td>
<td>0.427</td>
<td>0.880</td>
<td>0.192</td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>1.340</td>
<td>0.309</td>
<td>0.392</td>
<td>0.885</td>
<td>0.192</td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/20</td>
<td>1.286</td>
<td>0.305</td>
<td>0.456</td>
<td>0.913</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>1.349</td>
<td>0.311</td>
<td>0.471</td>
<td>0.913</td>
<td>0.186</td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>1.271</td>
<td>0.294</td>
<td>0.513</td>
<td>0.853</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>1.343</td>
<td>0.313</td>
<td>0.516</td>
<td>0.844</td>
<td>0.190</td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>1.271</td>
<td>0.306</td>
<td>0.427</td>
<td>0.880</td>
<td>0.192</td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>1.340</td>
<td>0.309</td>
<td>0.392</td>
<td>0.885</td>
<td>0.192</td>
</tr>
<tr>
<td>DATA SETS</td>
<td>DESIGN</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unbiased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>0.049</td>
<td>1.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>0.033</td>
<td>1.044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>0.040</td>
<td>1.035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/60</td>
<td>0.050</td>
<td>1.076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>0.035</td>
<td>1.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>0.036</td>
<td>1.035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>0.040</td>
<td>0.986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT8</td>
<td>500/40/40</td>
<td>0.025</td>
<td>0.990</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/20</td>
<td>0.030</td>
<td>0.962</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>0.010</td>
<td>0.711</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/20</td>
<td>0.033</td>
<td>0.957</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>0.011</td>
<td>0.965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/20</td>
<td>0.028</td>
<td>0.982</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>0.016</td>
<td>0.987</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>0.009</td>
<td>0.956</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>-0.010</td>
<td>0.965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>0.017</td>
<td>0.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>-0.004</td>
<td>0.963</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4

Repeated Measures Analysis of Variance on the Item Characteristic Curve

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (G)</td>
<td>0.079</td>
<td>1</td>
<td>0.079</td>
<td>0.423</td>
</tr>
<tr>
<td>Test Length (N)</td>
<td>0.613</td>
<td>2</td>
<td>0.306</td>
<td>1.629</td>
</tr>
<tr>
<td>Bias (I)</td>
<td>1.264</td>
<td>2</td>
<td>0.632</td>
<td>3.358*</td>
</tr>
<tr>
<td>GH</td>
<td>0.016</td>
<td>2</td>
<td>0.008</td>
<td>0.043</td>
</tr>
<tr>
<td>GI</td>
<td>0.104</td>
<td>2</td>
<td>0.052</td>
<td>0.277</td>
</tr>
<tr>
<td>HI</td>
<td>0.010</td>
<td>4</td>
<td>0.002</td>
<td>0.014</td>
</tr>
<tr>
<td>GHI</td>
<td>0.032</td>
<td>4</td>
<td>0.008</td>
<td>0.043</td>
</tr>
<tr>
<td>Error</td>
<td>245.067</td>
<td>1302</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>Trial (R)</td>
<td>0.000</td>
<td>1</td>
<td>0.000</td>
<td>0.988</td>
</tr>
<tr>
<td>RG</td>
<td>0.079</td>
<td>1</td>
<td>0.079</td>
<td>1.730</td>
</tr>
<tr>
<td>RH</td>
<td>0.222</td>
<td>2</td>
<td>0.111</td>
<td>2.408</td>
</tr>
<tr>
<td>RI</td>
<td>2.923</td>
<td>2</td>
<td>0.461</td>
<td>31.696*</td>
</tr>
<tr>
<td>RGH</td>
<td>0.016</td>
<td>2</td>
<td>0.008</td>
<td>0.176</td>
</tr>
<tr>
<td>RGI</td>
<td>0.104</td>
<td>2</td>
<td>0.052</td>
<td>1.130</td>
</tr>
<tr>
<td>RHI</td>
<td>0.014</td>
<td>4</td>
<td>0.003</td>
<td>0.076</td>
</tr>
<tr>
<td>RGH1</td>
<td>0.032</td>
<td>4</td>
<td>0.008</td>
<td>0.176</td>
</tr>
<tr>
<td>Error</td>
<td>60.054</td>
<td>1302</td>
<td>0.046</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 5

Repeated Measures Analysis of Variance on the bi item Characteristics Curve

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (G)</td>
<td>0.075</td>
<td>1</td>
<td>0.075</td>
<td>0.100</td>
</tr>
<tr>
<td>Test Length (N)</td>
<td>1.061</td>
<td>2</td>
<td>0.530</td>
<td>0.763</td>
</tr>
<tr>
<td>Bias (I)</td>
<td>1.159</td>
<td>2</td>
<td>0.579</td>
<td>0.768</td>
</tr>
<tr>
<td>GH</td>
<td>0.004</td>
<td>2</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>GI</td>
<td>0.013</td>
<td>2</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>HI</td>
<td>0.131</td>
<td>4</td>
<td>0.032</td>
<td>0.043</td>
</tr>
<tr>
<td>GHI</td>
<td>0.030</td>
<td>4</td>
<td>0.007</td>
<td>0.009</td>
</tr>
<tr>
<td>Error</td>
<td>982.204</td>
<td>1302</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>Trial (R)</td>
<td>4.514</td>
<td>1</td>
<td>4.514</td>
<td>180.150*</td>
</tr>
<tr>
<td>RG</td>
<td>0.075</td>
<td>1</td>
<td>0.075</td>
<td>3.019</td>
</tr>
<tr>
<td>RH</td>
<td>0.226</td>
<td>2</td>
<td>0.113</td>
<td>4.512</td>
</tr>
<tr>
<td>RI</td>
<td>1.532</td>
<td>2</td>
<td>0.766</td>
<td>30.585*</td>
</tr>
<tr>
<td>RGH</td>
<td>0.004</td>
<td>2</td>
<td>0.002</td>
<td>0.079</td>
</tr>
<tr>
<td>RGI</td>
<td>0.013</td>
<td>2</td>
<td>0.006</td>
<td>0.268</td>
</tr>
<tr>
<td>RHI</td>
<td>0.066</td>
<td>4</td>
<td>0.001</td>
<td>0.065</td>
</tr>
<tr>
<td>RGH1</td>
<td>0.030</td>
<td>4</td>
<td>0.007</td>
<td>0.300</td>
</tr>
<tr>
<td>Error</td>
<td>32.624</td>
<td>1302</td>
<td>0.025</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (G)</td>
<td>0.017</td>
<td>1</td>
<td>0.017</td>
<td>3.638</td>
</tr>
<tr>
<td>Test Length (H)</td>
<td>0.040</td>
<td>2</td>
<td>0.020</td>
<td>4.231*</td>
</tr>
<tr>
<td>Bias (I)</td>
<td>0.028</td>
<td>2</td>
<td>0.014</td>
<td>3.050*</td>
</tr>
<tr>
<td>GH</td>
<td>0.003</td>
<td>2</td>
<td>0.001</td>
<td>0.349</td>
</tr>
<tr>
<td>GI</td>
<td>0.000</td>
<td>2</td>
<td>0.000</td>
<td>0.104</td>
</tr>
<tr>
<td>HY</td>
<td>0.008</td>
<td>4</td>
<td>0.002</td>
<td>3.434</td>
</tr>
<tr>
<td>GHI</td>
<td>0.001</td>
<td>4</td>
<td>0.000</td>
<td>0.079</td>
</tr>
<tr>
<td>Error</td>
<td>6.180</td>
<td>1302</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Trial (S)</td>
<td>0.189</td>
<td>1</td>
<td>0.189</td>
<td>64.239*</td>
</tr>
<tr>
<td>RG</td>
<td>0.017</td>
<td>1</td>
<td>0.017</td>
<td>5.852*</td>
</tr>
<tr>
<td>RH</td>
<td>0.226</td>
<td>2</td>
<td>0.113</td>
<td>4.512</td>
</tr>
<tr>
<td>RY</td>
<td>0.028</td>
<td>2</td>
<td>0.014</td>
<td>4.906*</td>
</tr>
<tr>
<td>RGH</td>
<td>0.003</td>
<td>2</td>
<td>0.001</td>
<td>0.561</td>
</tr>
<tr>
<td>RGI</td>
<td>0.000</td>
<td>2</td>
<td>0.000</td>
<td>0.167</td>
</tr>
<tr>
<td>RHI</td>
<td>0.008</td>
<td>4</td>
<td>0.002</td>
<td>0.699</td>
</tr>
<tr>
<td>RGHI</td>
<td>0.001</td>
<td>4</td>
<td>0.000</td>
<td>0.127</td>
</tr>
<tr>
<td>Error</td>
<td>3.842</td>
<td>1302</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 7  
Repeated Measures Analysis of Variance on the Person Ability Parameter (Bv)

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size (G)</td>
<td>16.133</td>
<td>1</td>
<td>16.133</td>
<td>8.771*</td>
</tr>
<tr>
<td>Test Length (H)</td>
<td>19.193</td>
<td>2</td>
<td>9.596</td>
<td>5.217*</td>
</tr>
<tr>
<td>Bias (I)</td>
<td>15.085</td>
<td>2</td>
<td>7.542</td>
<td>4.101*</td>
</tr>
<tr>
<td>GH</td>
<td>23.314</td>
<td>2</td>
<td>11.657</td>
<td>6.338*</td>
</tr>
<tr>
<td>GI</td>
<td>24.198</td>
<td>2</td>
<td>12.099</td>
<td>6.578*</td>
</tr>
<tr>
<td>NI</td>
<td>44.456</td>
<td>4</td>
<td>11.114</td>
<td>6.042*</td>
</tr>
<tr>
<td>GHI</td>
<td>45.817</td>
<td>4</td>
<td>11.454</td>
<td>6.227*</td>
</tr>
<tr>
<td>Error</td>
<td>24796.800</td>
<td>13482</td>
<td>1.839</td>
<td></td>
</tr>
<tr>
<td>Trial (K)</td>
<td>1.442</td>
<td>1</td>
<td>1.442</td>
<td>49.104*</td>
</tr>
<tr>
<td>EG</td>
<td>0.059</td>
<td>1</td>
<td>0.059</td>
<td>2.010*</td>
</tr>
<tr>
<td>RH</td>
<td>0.224</td>
<td>2</td>
<td>0.112</td>
<td>3.826*</td>
</tr>
<tr>
<td>RI</td>
<td>0.929</td>
<td>2</td>
<td>0.464</td>
<td>15.820*</td>
</tr>
<tr>
<td>RGH</td>
<td>0.047</td>
<td>2</td>
<td>0.023</td>
<td>0.813</td>
</tr>
<tr>
<td>RGI</td>
<td>0.072</td>
<td>2</td>
<td>0.036</td>
<td>1.227</td>
</tr>
<tr>
<td>RH1</td>
<td>0.027</td>
<td>4</td>
<td>0.006</td>
<td>0.233</td>
</tr>
<tr>
<td>RGHI</td>
<td>0.064</td>
<td>4</td>
<td>0.016</td>
<td>0.550</td>
</tr>
<tr>
<td>Error</td>
<td>396.116</td>
<td>13482</td>
<td>0.029</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
<table>
<thead>
<tr>
<th>DATA SETS</th>
<th>DESIGN</th>
<th>ai</th>
<th>SD</th>
<th>( \hat{ai} )</th>
<th>Mean</th>
<th>SD</th>
<th>bi</th>
<th>SD</th>
<th>( \hat{bi} )</th>
<th>Mean</th>
<th>SD</th>
<th>ci</th>
<th>SD</th>
<th>( \hat{ci} )</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>1.223</td>
<td>0.262</td>
<td>1.313</td>
<td>0.399</td>
<td>0.481</td>
<td>0.902</td>
<td>-0.114</td>
<td>1.097</td>
<td>0.190</td>
<td>0.037</td>
<td>0.171</td>
<td>0.089</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>1.209</td>
<td>0.244</td>
<td>1.312</td>
<td>0.372</td>
<td>0.480</td>
<td>0.875</td>
<td>-0.115</td>
<td>1.001</td>
<td>0.190</td>
<td>0.037</td>
<td>0.181</td>
<td>0.081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>1.194</td>
<td>0.237</td>
<td>1.273</td>
<td>0.374</td>
<td>0.455</td>
<td>0.879</td>
<td>-0.101</td>
<td>1.004</td>
<td>0.192</td>
<td>0.038</td>
<td>0.180</td>
<td>0.074</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/40</td>
<td>1.221</td>
<td>0.262</td>
<td>1.369</td>
<td>0.355</td>
<td>0.481</td>
<td>0.902</td>
<td>-0.089</td>
<td>1.016</td>
<td>0.186</td>
<td>0.039</td>
<td>0.183</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>1.209</td>
<td>0.244</td>
<td>1.290</td>
<td>0.310</td>
<td>0.480</td>
<td>0.875</td>
<td>-0.109</td>
<td>0.995</td>
<td>0.190</td>
<td>0.027</td>
<td>0.183</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>1.194</td>
<td>0.237</td>
<td>1.283</td>
<td>0.317</td>
<td>0.455</td>
<td>0.879</td>
<td>-0.066</td>
<td>0.968</td>
<td>0.192</td>
<td>0.038</td>
<td>0.188</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>1.286</td>
<td>0.305</td>
<td>1.306</td>
<td>0.478</td>
<td>0.456</td>
<td>0.913</td>
<td>-0.138</td>
<td>1.202</td>
<td>0.186</td>
<td>0.039</td>
<td>0.152</td>
<td>0.096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT8</td>
<td>500/60/40</td>
<td>1.349</td>
<td>0.311</td>
<td>1.325</td>
<td>0.668</td>
<td>0.471</td>
<td>0.912</td>
<td>-0.131</td>
<td>1.214</td>
<td>0.186</td>
<td>0.039</td>
<td>0.153</td>
<td>0.093</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/80</td>
<td>1.271</td>
<td>0.294</td>
<td>1.259</td>
<td>0.423</td>
<td>0.513</td>
<td>0.853</td>
<td>-0.130</td>
<td>1.141</td>
<td>0.190</td>
<td>0.037</td>
<td>0.171</td>
<td>0.075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>1.343</td>
<td>0.313</td>
<td>1.293</td>
<td>0.434</td>
<td>0.516</td>
<td>0.846</td>
<td>-0.105</td>
<td>1.130</td>
<td>0.190</td>
<td>0.037</td>
<td>0.170</td>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/40</td>
<td>1.271</td>
<td>0.306</td>
<td>1.237</td>
<td>0.432</td>
<td>0.427</td>
<td>0.880</td>
<td>-0.090</td>
<td>1.146</td>
<td>0.192</td>
<td>0.038</td>
<td>0.160</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>1.340</td>
<td>0.309</td>
<td>1.271</td>
<td>0.434</td>
<td>0.392</td>
<td>0.885</td>
<td>-0.037</td>
<td>1.118</td>
<td>0.192</td>
<td>0.038</td>
<td>0.170</td>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/40</td>
<td>1.286</td>
<td>0.305</td>
<td>1.267</td>
<td>0.413</td>
<td>0.456</td>
<td>0.913</td>
<td>-0.105</td>
<td>1.202</td>
<td>0.186</td>
<td>0.039</td>
<td>0.165</td>
<td>0.098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>1.349</td>
<td>0.311</td>
<td>1.288</td>
<td>0.416</td>
<td>0.471</td>
<td>0.913</td>
<td>-0.100</td>
<td>1.202</td>
<td>0.186</td>
<td>0.039</td>
<td>0.160</td>
<td>0.102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>1.271</td>
<td>0.294</td>
<td>1.219</td>
<td>0.367</td>
<td>0.513</td>
<td>0.853</td>
<td>-0.098</td>
<td>1.113</td>
<td>0.190</td>
<td>0.037</td>
<td>0.176</td>
<td>0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>1.343</td>
<td>0.313</td>
<td>1.256</td>
<td>0.377</td>
<td>0.516</td>
<td>0.844</td>
<td>-0.074</td>
<td>1.100</td>
<td>0.190</td>
<td>0.037</td>
<td>0.178</td>
<td>0.066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>1.271</td>
<td>0.306</td>
<td>1.190</td>
<td>0.359</td>
<td>0.427</td>
<td>0.880</td>
<td>-0.071</td>
<td>1.106</td>
<td>0.192</td>
<td>0.038</td>
<td>0.177</td>
<td>0.069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>1.340</td>
<td>0.309</td>
<td>1.213</td>
<td>0.361</td>
<td>0.392</td>
<td>0.885</td>
<td>-0.036</td>
<td>1.089</td>
<td>0.192</td>
<td>0.038</td>
<td>0.179</td>
<td>0.070</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 9

**Means and Standard Deviations of the Distributions of the Known (by) and Estimated (by) Person Ability Parameters**

<table>
<thead>
<tr>
<th>DATA SETS</th>
<th>DESIGN</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unbiased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>0.049</td>
<td>1.068</td>
<td>0.042</td>
<td>1.004</td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>0.033</td>
<td>1.044</td>
<td>0.042</td>
<td>1.004</td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>0.040</td>
<td>1.035</td>
<td>0.042</td>
<td>1.003</td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/40</td>
<td>0.050</td>
<td>1.076</td>
<td>0.039</td>
<td>1.020</td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>0.035</td>
<td>1.041</td>
<td>0.039</td>
<td>1.020</td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>0.036</td>
<td>1.035</td>
<td>0.039</td>
<td>1.019</td>
</tr>
<tr>
<td><strong>Biased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>0.040</td>
<td>0.986</td>
<td>0.046</td>
<td>0.903</td>
</tr>
<tr>
<td>DAT8</td>
<td>500/40/40</td>
<td>0.025</td>
<td>0.990</td>
<td>0.046</td>
<td>0.903</td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/20</td>
<td>0.030</td>
<td>0.962</td>
<td>0.046</td>
<td>0.903</td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>0.010</td>
<td>0.971</td>
<td>0.046</td>
<td>0.903</td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/20</td>
<td>0.031</td>
<td>0.957</td>
<td>0.046</td>
<td>0.899</td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>0.011</td>
<td>0.965</td>
<td>0.046</td>
<td>0.899</td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/20</td>
<td>0.028</td>
<td>0.982</td>
<td>0.038</td>
<td>0.909</td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>0.016</td>
<td>0.987</td>
<td>0.038</td>
<td>0.909</td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>0.009</td>
<td>0.956</td>
<td>0.038</td>
<td>0.909</td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>-0.010</td>
<td>0.965</td>
<td>0.038</td>
<td>0.909</td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>0.017</td>
<td>0.954</td>
<td>0.038</td>
<td>0.907</td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>-0.004</td>
<td>0.963</td>
<td>0.038</td>
<td>0.907</td>
</tr>
</tbody>
</table>
Table 10
Average Absolute Differences and Root Mean Square Errors for Each Item Characteristic Curve Parameter for 18 Data Sets

<table>
<thead>
<tr>
<th>DATA SETS</th>
<th>DESIGN</th>
<th>$\hat{a}_i$</th>
<th>AAD</th>
<th>RMSE</th>
<th>$\hat{b}_i$</th>
<th>AAD</th>
<th>RMSE</th>
<th>$\hat{c}_i$</th>
<th>AAD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unbiased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>0.237</td>
<td>0.313</td>
<td>0.681</td>
<td>1.373</td>
<td>0.067</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>0.224</td>
<td>0.299</td>
<td>0.664</td>
<td>1.356</td>
<td>0.059</td>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>0.222</td>
<td>0.291</td>
<td>0.613</td>
<td>1.300</td>
<td>0.056</td>
<td>0.074</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/40</td>
<td>0.179</td>
<td>0.249</td>
<td>0.607</td>
<td>1.310</td>
<td>0.038</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>0.150</td>
<td>0.206</td>
<td>0.635</td>
<td>1.346</td>
<td>0.041</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>0.155</td>
<td>0.208</td>
<td>0.567</td>
<td>1.261</td>
<td>0.043</td>
<td>0.057</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Biased</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>0.321</td>
<td>0.403</td>
<td>0.693</td>
<td>1.364</td>
<td>0.075</td>
<td>0.097</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT8</td>
<td>500/40/40</td>
<td>0.329</td>
<td>0.408</td>
<td>0.701</td>
<td>1.384</td>
<td>0.074</td>
<td>0.092</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/20</td>
<td>0.285</td>
<td>0.360</td>
<td>0.736</td>
<td>1.435</td>
<td>0.060</td>
<td>0.076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>0.287</td>
<td>0.350</td>
<td>0.705</td>
<td>1.377</td>
<td>0.061</td>
<td>0.079</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/20</td>
<td>0.298</td>
<td>0.366</td>
<td>0.636</td>
<td>1.288</td>
<td>0.080</td>
<td>0.098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>0.302</td>
<td>0.371</td>
<td>0.532</td>
<td>1.138</td>
<td>0.081</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/20</td>
<td>0.238</td>
<td>0.300</td>
<td>0.687</td>
<td>1.374</td>
<td>0.075</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>0.265</td>
<td>0.325</td>
<td>0.687</td>
<td>1.374</td>
<td>0.079</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>0.218</td>
<td>0.290</td>
<td>0.700</td>
<td>1.405</td>
<td>0.069</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>0.219</td>
<td>0.280</td>
<td>0.664</td>
<td>1.360</td>
<td>0.049</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>0.217</td>
<td>0.277</td>
<td>0.579</td>
<td>1.271</td>
<td>0.051</td>
<td>0.066</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>0.244</td>
<td>0.301</td>
<td>0.504</td>
<td>1.142</td>
<td>0.052</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 11

Average Absolute Differences and Root Mean Square Errors for the Person Ability Score for 18 Data Sets

<table>
<thead>
<tr>
<th>DATA SETS</th>
<th>DESIGN</th>
<th>AAD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiased</td>
<td>DAT1</td>
<td>500/40</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td>DAT2</td>
<td>500/80</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>DAT3</td>
<td>500/100</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>DAT4</td>
<td>1000/40</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>DAT5</td>
<td>1000/80</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>DAT6</td>
<td>1000/100</td>
<td>0.156</td>
</tr>
<tr>
<td>Biased</td>
<td>DAT7</td>
<td>500/40/20</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>DAT8</td>
<td>500/40/40</td>
<td>0.231</td>
</tr>
<tr>
<td></td>
<td>DAT9</td>
<td>500/80/20</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td>DAT10</td>
<td>500/80/40</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>DAT11</td>
<td>500/100/20</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>DAT12</td>
<td>500/100/40</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>DAT13</td>
<td>1000/40/20</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>DAT14</td>
<td>1000/40/40</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>DAT15</td>
<td>1000/80/20</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>DAT16</td>
<td>1000/80/40</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>DAT17</td>
<td>1000/100/20</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>DAT18</td>
<td>1000/100/40</td>
<td>0.159</td>
</tr>
</tbody>
</table>
Table 12

Simple Regressions and Pearson Product-Moment Correlations of the Item Characteristic Curve Parameter Estimates ($a_i$, $b_i$ and $c_i$) on the Known Parameters ($a_i$, $b_i$ and $c_i$)

<table>
<thead>
<tr>
<th>DATA SETS/</th>
<th>PARAMETERS</th>
<th>DESIGN</th>
<th>$r$</th>
<th>$r^2$</th>
<th>SEE</th>
<th>Intercept ($a$)</th>
<th>Slope ($b$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAT1 500/40</td>
<td>$a_i.a_i$</td>
<td>0.649</td>
<td>0.421</td>
<td>0.020</td>
<td>0.661</td>
<td>0.425</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.225</td>
<td>0.050</td>
<td>0.890</td>
<td>0.501</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.151</td>
<td>0.023</td>
<td>0.039</td>
<td>0.174</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>PAT2 500/80</td>
<td>$a_i.a_i$</td>
<td>0.650</td>
<td>0.423</td>
<td>0.186</td>
<td>0.648</td>
<td>0.427</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.149</td>
<td>0.022</td>
<td>0.870</td>
<td>0.494</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.302</td>
<td>0.091</td>
<td>0.035</td>
<td>0.164</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td>PAT3 500/100</td>
<td>$a_i.a_i$</td>
<td>0.655</td>
<td>0.430</td>
<td>0.150</td>
<td>0.664</td>
<td>0.415</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.218</td>
<td>0.047</td>
<td>0.862</td>
<td>0.474</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.312</td>
<td>0.097</td>
<td>0.036</td>
<td>0.164</td>
<td>0.156</td>
<td></td>
</tr>
<tr>
<td>PAT4 1000/40</td>
<td>$a_i.a_i$</td>
<td>0.825</td>
<td>0.681</td>
<td>0.159</td>
<td>0.385</td>
<td>0.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.128</td>
<td>0.051</td>
<td>0.890</td>
<td>0.498</td>
<td>0.202</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.613</td>
<td>0.376</td>
<td>0.031</td>
<td>0.118</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>PAT5 1000/80</td>
<td>$a_i.a_i$</td>
<td>0.785</td>
<td>0.617</td>
<td>0.152</td>
<td>0.610</td>
<td>0.619</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.155</td>
<td>0.024</td>
<td>0.859</td>
<td>0.454</td>
<td>0.136</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.489</td>
<td>0.239</td>
<td>0.032</td>
<td>0.138</td>
<td>0.281</td>
<td></td>
</tr>
<tr>
<td>PAT6 1000/100</td>
<td>$a_i.a_i$</td>
<td>0.203</td>
<td>0.645</td>
<td>0.142</td>
<td>0.423</td>
<td>0.601</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.221</td>
<td>0.049</td>
<td>0.851</td>
<td>0.468</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.454</td>
<td>0.206</td>
<td>0.034</td>
<td>0.142</td>
<td>0.272</td>
<td></td>
</tr>
<tr>
<td>PAT7 500/40/20</td>
<td>$a_i.a_i$</td>
<td>0.533</td>
<td>0.292</td>
<td>0.261</td>
<td>0.843</td>
<td>0.530</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.332</td>
<td>0.110</td>
<td>0.872</td>
<td>0.490</td>
<td>0.252</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.305</td>
<td>0.093</td>
<td>0.037</td>
<td>0.167</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td>PAT8 500/40/48</td>
<td>$a_i.a_i$</td>
<td>0.498</td>
<td>0.248</td>
<td>0.273</td>
<td>0.009</td>
<td>0.331</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.321</td>
<td>0.103</td>
<td>0.576</td>
<td>0.502</td>
<td>0.241</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.356</td>
<td>0.127</td>
<td>0.037</td>
<td>0.183</td>
<td>0.183</td>
<td></td>
</tr>
<tr>
<td>PAT9 500/80/20</td>
<td>$a_i.a_i$</td>
<td>0.538</td>
<td>0.290</td>
<td>0.369</td>
<td>0.799</td>
<td>0.374</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$b_i.b_i$</td>
<td>0.185</td>
<td>0.034</td>
<td>0.843</td>
<td>0.531</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$c_i.c_i$</td>
<td>0.247</td>
<td>0.061</td>
<td>0.036</td>
<td>0.168</td>
<td>0.122</td>
<td></td>
</tr>
</tbody>
</table>

(Table 12 continues)
<table>
<thead>
<tr>
<th>DATE SETS/ PARAMETERS CORRELATED</th>
<th>Intercept Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DESIGN r  r²  SEE (a)  (b)</td>
</tr>
<tr>
<td>D AT 10 500/80/40</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.567  0.369  0.250  0.778  0.437</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.240  0.057  0.024  0.534  0.179</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.260  0.067  0.036  0.168  0.123</td>
</tr>
<tr>
<td>D AT 11 500/100/20</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.617  0.451  0.229  0.642  0.574</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.434  0.188  0.801  0.404  0.352</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.352  0.124  0.036  0.158  0.192</td>
</tr>
<tr>
<td>D AT 12 500/100/40</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.535  0.308  0.258  0.638  0.394</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.460  0.211  0.789  0.405  0.364</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.188  0.035  0.037  0.108  0.072</td>
</tr>
<tr>
<td>D AT 13 1000/40/20</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.579  0.461  0.226  0.651  0.501</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.302  0.091  0.061  0.480  0.229</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.140  0.019  0.095  0.217  0.056</td>
</tr>
<tr>
<td>D AT 14 1000/40/40</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.638  0.407  0.242  0.724  0.476</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.307  0.094  0.680  0.493  0.232</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.154  0.015  0.039  0.178  0.048</td>
</tr>
<tr>
<td>D AT 15 1000/80/20</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.642  0.412  0.226  0.644  0.513</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.181  0.032  0.644  0.526  0.130</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.015  0.172  0.032  0.159  0.128</td>
</tr>
<tr>
<td>D AT 16 1000/80/40</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.712  0.507  0.221  0.599  0.391</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.315  0.014  0.829  0.526  0.165</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.427  0.102  0.037  0.149  0.227</td>
</tr>
<tr>
<td>D AT 17 1000/100/20</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.684  0.474  0.222  0.573  0.585</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.315  0.100  0.839  0.644  0.252</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.367  0.135  0.035  0.156  0.205</td>
</tr>
<tr>
<td>D AT 18 1000/100/40</td>
<td></td>
</tr>
<tr>
<td>ai ai</td>
<td>0.671  0.451  0.229  0.642  0.574</td>
</tr>
<tr>
<td>bi bi</td>
<td>0.456  0.158  0.801  0.504  0.352</td>
</tr>
<tr>
<td>ci ci</td>
<td>0.357  0.124  0.036  0.158  0.192</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 13

Simple Regressions and Pearson Product-Moment Correlations of the Person Ability Parameter Estimates (Bv) on the Known Person Ability Scores (Bv)

<table>
<thead>
<tr>
<th>DATA SETS</th>
<th>DESIGN</th>
<th>r</th>
<th>r²</th>
<th>SEE</th>
<th>Intercept (a)</th>
<th>Slope (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT1</td>
<td>500/40</td>
<td>0.957</td>
<td>0.915</td>
<td>0.310</td>
<td>0.006</td>
<td>1.109</td>
</tr>
<tr>
<td>DAT2</td>
<td>500/80</td>
<td>0.978</td>
<td>0.957</td>
<td>0.216</td>
<td>-0.009</td>
<td>1.018</td>
</tr>
<tr>
<td>DAT3</td>
<td>500/100</td>
<td>0.981</td>
<td>0.963</td>
<td>0.199</td>
<td>-0.002</td>
<td>1.013</td>
</tr>
<tr>
<td>DAT4</td>
<td>1000/40</td>
<td>0.959</td>
<td>0.921</td>
<td>0.302</td>
<td>-0.010</td>
<td>1.014</td>
</tr>
<tr>
<td>DAT5</td>
<td>1000/80</td>
<td>0.977</td>
<td>0.955</td>
<td>0.218</td>
<td>0.003</td>
<td>0.999</td>
</tr>
<tr>
<td>DAT6</td>
<td>1000/100</td>
<td>0.980</td>
<td>0.961</td>
<td>0.202</td>
<td>-0.002</td>
<td>0.996</td>
</tr>
<tr>
<td>Biased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT7</td>
<td>500/40/20</td>
<td>0.954</td>
<td>0.911</td>
<td>0.295</td>
<td>-0.007</td>
<td>1.046</td>
</tr>
<tr>
<td>DAT8</td>
<td>500/40/40</td>
<td>0.955</td>
<td>0.912</td>
<td>0.294</td>
<td>0.022</td>
<td>1.051</td>
</tr>
<tr>
<td>DAT9</td>
<td>500/80/20</td>
<td>0.978</td>
<td>0.958</td>
<td>0.198</td>
<td>-0.017</td>
<td>1.046</td>
</tr>
<tr>
<td>DAT10</td>
<td>500/80/40</td>
<td>0.978</td>
<td>0.962</td>
<td>0.186</td>
<td>-0.038</td>
<td>1.056</td>
</tr>
<tr>
<td>DAT11</td>
<td>500/100/20</td>
<td>0.980</td>
<td>0.962</td>
<td>0.186</td>
<td>-0.016</td>
<td>1.043</td>
</tr>
<tr>
<td>DAT12</td>
<td>500/100/40</td>
<td>0.980</td>
<td>0.961</td>
<td>0.189</td>
<td>-0.037</td>
<td>1.051</td>
</tr>
<tr>
<td>DAT13</td>
<td>1000/40/20</td>
<td>0.952</td>
<td>0.907</td>
<td>0.299</td>
<td>-0.011</td>
<td>1.031</td>
</tr>
<tr>
<td>DAT14</td>
<td>1000/40/40</td>
<td>0.953</td>
<td>0.908</td>
<td>0.299</td>
<td>-0.022</td>
<td>1.036</td>
</tr>
<tr>
<td>DAT15</td>
<td>1000/80/20</td>
<td>0.975</td>
<td>0.952</td>
<td>0.209</td>
<td>-0.030</td>
<td>1.028</td>
</tr>
<tr>
<td>DAT16</td>
<td>1000/80/40</td>
<td>0.974</td>
<td>0.950</td>
<td>0.215</td>
<td>-0.049</td>
<td>1.036</td>
</tr>
<tr>
<td>DAT17</td>
<td>1000/100/20</td>
<td>0.978</td>
<td>0.958</td>
<td>0.195</td>
<td>-0.021</td>
<td>1.028</td>
</tr>
<tr>
<td>DAT18</td>
<td>1000/100/40</td>
<td>0.978</td>
<td>0.957</td>
<td>0.199</td>
<td>-0.043</td>
<td>1.038</td>
</tr>
</tbody>
</table>
Figure 1. The Person Response Curve, adapted from Wright, B.D. & Stone, M.H. (1979). Best Test Design: Rasch Measurement, Chicago: MESA Press.
Figure 2. The Item Characteristic Curve.
Figure 3. A Bar Graph of the Average Absolute Differences for the $a_i$, $b_i$, and $c_i$ item characteristic curve parameters on tests which were either 0, 20, 40 percent biased.
A Bar Graph of the Average Absolute Differences for Bv on tests which were either 0, 20, or 40 percent biased.

Figure 4.

AVERAGE ABSOLUTE DIFFERENCES (LOGITS)

<table>
<thead>
<tr>
<th>BIAS (PERCENT)</th>
<th>0</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.1  0.2  0.3  0.4  0.5
APPENDIX A

ALGEBRAIC PROCEDURES
FOR THE
ONE-PARAMETER ITEM RESPONSE MODEL

Adapted from Wright, B. D. & Stone, M. H. (1979).
Best Test Design: Rasch Measurement, Chicago: MESA Press
The algebraic procedures for the one-parameter model, PROX (Cohen, 1976), are reproduced from Wright and Stone (1979, p. 28-45). These procedures approximate results obtained with the more elaborate and hence, the more accurate maximum likelihood procedures employed by other one-parameter models (cf., Wright & Stone, 1979). Still, PROX achieves the basic aims of item-response, item analysis, namely, the linearization of the latent-variable continuum and adjustment of parameter estimates for the local affects of item calibration and sample ability.

PROX adequately accounts for these affects by just a mean and standard deviation, and these procedures do so with sample sizes as small as 500 examinees. The data used for illustration of PROX comes from the administration of the 18-item KNOX Cube Test, a subtest of the Arthur Point Scale (Arthur, 1947) to 35 students (cf., Wright & Stone, 1979, p. 28). These data illustrate the basic tenets of item response theory, and present a method for performing item response item analysis which is applicable to data obtained in industrial-organizational research.

Table 1 shows the responses of 35 students to a single administration of the 18 item Knox Cube Test. The data are arranged in a person-item matrix. The 18 items are listed across the top of the table. The 35 persons on which the data were obtained are listed in the first left column. A correct response is recorded as a 1, and an incorrect response is recorded as a 0. The number of items correctly answered by each examinee, that is, the person
scores, are given at the end of each row in the last column on the right. The number of examinees which correctly answered each item, that is, the item scores, are given at the bottom of each column.

------------------------

Insert Table 1 about here
------------------------

The general plan for accomplishing item response item analysis with the PROX method begins with the editing of the person-item data matrix presented in Table 1. Persons and items for which no definite estimates of ability or difficulty exist are removed from the unedited person-item matrix. Persons who passed or missed all test items, and items passed or missed by all examinees, are removed from the person-item matrix.

The boundary lines drawn in Table 1 show the items and persons removed by the editing process. Items 1, 2, and 3 were removed because they were answered correctly by all examinees. Removing these three items, brought about the removal of Person 35 because this person had only these three items correct and hence, none correct after items 1, 2 and 3 were removed. Item 18 was removed because no persons answered this item correctly.

Editing a data matrix may require several such cycles because removing items can necessitate removing persons and vice versa. For example, had there been a person who had succeeded on all but Item 18, then removal of Item 18 would have left this person with a perfect score on the remaining items and so, that person would also
have had to be removed. The item-person matrix is edited until only those items and persons which provide information about person abilities and item difficulties, that is the items and persons with p-values between but not including 0.0 and 1.0 are included.

The edited and ordered person-item matrix is presented in Table 2. It is arranged so that person scores are ordered from low to high with regard to their respective p-values. The person p-values are given in the right-most column. The item scores are ordered from high to low with regard to their respective p-values, and these are given in the bottom-most row.

---

Insert Table 2 about here

---

From the edited and ordered person-item data matrix presented in Table 2, a grouped distribution of 10 different item scores and their respective logits incorrect is obtained. Also, the mean and variance of the distribution of item logits for the 14 items is computed. These procedures are described in Table 3.

---

Insert Table 3 about here

---

The first column in Table 3 identifies the item score group index (i). There are 10 individual item groups identified in Table 3. These groups are listed from the easiest to most difficult groups of test items or item subsets (see column 3). Column 2
identifies the names of the items which comprise each group. For example, Item Score Group 2 is comprised of two items, items 5 and 7. Both of these items are relatively easy items; they were passed by 31 examinees and have p-values of .91. Item Score Group 10 is comprised of three items, items 15, 16 and 17. Except for item 18, which was removed from the person-item matrix during the editing process because no examinees passed this item, these were the most difficult items included in the 18-item KNOX Cube Test. Items 15, 16 and 17 were passed by a single examinee and have p-values of .03.

The item scores (Si), that is, the number of examinees which correctly answered the individual test items, are given in column 3. Column 4 identifies the frequency (fi) or number of items in each Item Score Group. For example, item 4 was in the first score group and had a score of 32. Items 5 and 7 were in the second group.

The proportion correct values (pi), that is, the number of examinees which correctly answered the item (Si) out of the total sample (N), are presented in column 5. Si is the item score value presented in column 3 and the total sample (N) is the number of examinees which remained in the edited person-item matrix. In the present example, examinee number 35 was removed from this matrix, so that the resultant sample size (N) was 34 and not 35.

Column 6 converts the proportion correct values (pi) obtained in column 5 to values for the proportion incorrect (1 - pi). Log conversions of the values in column 6 yield logit incorrect values.
(Xi). These are listed in column 7. The logit incorrect values (Xi) are multiplied by the item score group frequencies (fi) to obtain the values reported in Column 8 (fixi). Column 9 squares the logit incorrect values (Xi) reported in column 7 and multiplies these squared values by their respective item frequencies (fi), to produce frequency times LOGIT squared values (fixi^2).

Before discussing the values provided in column 10 of Table 3, there are three additional values which are obtained through the summation of columns 4, 8 and 9 which need to be described. Each of these summations is performed across the item score groups listed in column 1; they sum across groups 1 to 10. The first value (Σfi) sums the item frequencies (fi) listed in column 4, across the item score groups (i). It defines the number of items included in the edited person-item data matrix, and in the present example 14 items remained after the editing process.

The frequency X LOGIT (fixi) values are summated across the 10 item score groups to obtain the value (Σfixi). This value is given under column 8. Under column 9, the value (Σfixi^2) is provided. This value is obtained by summing the frequency (fi) times LOGIT squared values (Xi^2) listed in column 9. The three values, (Σfi), (Σfixi) and (Σfixi^2), obtained through the summation of the values listed in columns 4, 8 and 9, are necessary for the computation of the mean (Xi) and variance (Ui) of the distribution of item logit incorrect values reported in column 7. Now, we can begin to describe the computation of the values given in column 10 of Table 3.
Equation 1 presents the algebraic formulations for arriving at the means \( \bar{X}_i \) of the distribution of item logit incorrect values \( X_i \). \( \bar{X}_i \) is obtained by dividing the sum of item frequency times the item logit incorrect values \( \sum f_i x_i \) by the number of items included in the edited person-item matrix \( f_i \).

\[
\bar{X}_i = \frac{1}{\sum f_i} \sum f_i x_i \sum f_i
\]

Equation 1

\( \bar{X}_i \) is the mean of the distribution of item logit incorrect values,

\( \sum f_i x_i \) is the sum of the item frequency \( f_i \) times item logit incorrect scores \( X_i \), and

\( \sum f_i \) is the item frequency or number of items in each item score group.

Equation 2 presents the algebraic computations for obtaining the variance \( \sigma^2_i \) of the distribution of item logit incorrect values \( X_i \). \( \sigma^2_i \) is obtained by squaring the mean item logit incorrect \( \bar{X}_i^2 \) and multiplying the obtained value by the item score group frequency \( f_i \). These values are summated across the different item score groups to yield \( \sum f_i x_i^2 \). This value \( \sum f_i x_i^2 \) is then subtracted from \( \sum f_i x_i^2 \), which is the sum of the item frequencies \( f_i \) times the squared incorrect item logit listed in column 9. The obtained value is then divided by the number of items included in the edited person-item matrix minus 1, that is, \( \sum f_i - 1 \).
\[ U_i = \frac{\sum_{i} f_i x_i^2 - \sum_{i} f_i x_i \cdot i^2}{\sum_{i} f_i - 1} \]  \hspace{1cm} \text{Equation 2}

\[ \sum_{i} f_i x_i^2 \] is the sum of the products of the item frequency \((f_i)\) and the squared item logit incorrect \((x_i)\) values across the item score groups,

\[ \sum_{i} f_i x_i \cdot i^2 \] is the sum of the products of the item frequency \((f_i)\) values and the square of the mean item logit incorrect \((x_i)\), and

\[ \sum_{i} f_i - 1 \] is the item frequency or number of items in the edited person-item matrix minus one.

The equations for deriving the mean and variance of the distributions of item logit incorrect values are provided at the bottom of Table 3, as are examples which use the values listed in Table 3. The mean item logit incorrect value \((X_i)\) is used in column 10 to arrive at the initial item calibration \((b_i)\). The values listed in column 10 are obtained by subtracting the mean item logit incorrect value \((X_i)\) from the item logit incorrect value \((X_i)\), that is, \((X_i - X_i)\), for each item score group \((i)\). This yields the initial item calibrations \((b_i)\).

Table 4 describes a similar set of procedures for arriving at person ability score estimates. The values presented in this Table are also based on the edited and ordered person-item matrix presented in Table 2. The algebraic computations described in Table 4 arrive at values for both the mean and variance of the distributions of person logit correct scores.

Insert Table 4 about here

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The first column in Table 4 presents the person ability score group index \((v)\). In actuality, it is redundant with column 2 which presents the possible person ability scores \((S_v)\); there is one group \((v)\) for each of the possible 13 ability scores \((S_v)\). \(S_v\) is the number of items answered correctly by each examinee, that is, the ability score. Note that \(14\) is not a possible person ability score. Remember that if a person correctly answers all items, the data obtained on that examinee is removed from the item-person matrix. If we let \(n\) be the number of items included in the edited person-item matrix, then \(S_v\), or the number of possible person ability scores, will be \((n-1)\). Unlike Table 3, the person names for each person ability score group are not identified. However, plots of the person ability score distributions are easily obtained (cf., Wright & Stone, 1979).

Column 3 shows the person ability score frequencies \((f_v)\). These values are the number of persons who obtained each possible ability score. The proportion correct \((p_v)\) is given in column 4. It is the number of items correctly answered \((S_v)\) out of the total number of items included in the edited item-person matrix \((n)\). The proportion incorrect \((1 - p_v)\) is given in column 5.

The person LOGIT correct \((X_v)\) is given in column 6. The person LOGIT correct is a logarithmic conversion of the proportion correct \((p_v)\) divided by the proportion incorrect \((1 - p_v)\), so that, \(X_v = \ln \left( \frac{p_v}{1 - p_v} \right)\). Column 7 takes the person ability score frequency \((f_v)\) and multiplies it by the person LOGIT correct \((X_v)\).
values. This is done for all possible person ability scores. Similarly, column 8 multiplies the person ability score frequency \((fv)\) shown in column 2 and the square of the person logit correct values \((x_v^2)\) for all possible person ability score groups \((v)\).

Before discussing column 9 of Table 4, there are three additional computations obtained through the summation of columns 3, 7 and 8 which need to be described. Each of these summations is performed across all person ability score groups \((v)\); they sum across groups 1 to 13. The first value, \((\sum fv)\), sums the person frequencies \((fv)\) at each of the 13 ability score group \((V)\). The value \((\sum fv)\) describes the number of persons included in the edited person-item matrix; in the present example, 34 persons remained in the edited matrix. The second computation sums the values \((fvX_v)\) shown in column 7 across the 13 person ability score groups. The third value results from the summation of the values \((fvX_v)^2\) shown in column 8. As with the computation of the mean and variance values for the distribution of item logits incorrect, the values obtained from the summation of columns 3, 7 and 9 are also used in the algebraic computation of the mean and variance of the distribution of person logits correct.

Equation 3 presents the algebraic computations for arriving at the mean \((x_r)\) of the mean of the distributions of person logit correct scores \((Sv)\). \(x_r\) is obtained by dividing the sum of the person frequency times the person logit correct values \((fvX_v)\) by the total number of persons included in the edited item-person matrix \((\sum fv)\).
\[
X. = \frac{\sum_{i=1}^{V} f_{v}X_{v}}{\sum_{i=1}^{V} f_{v}} \tag{Equation 3}
\]

\(X.\) is the mean of the distribution of person logit correct scores, \\
\(f_{v}X_{v}\) is the sum of the person frequency times the person logit correct scores, and \\
\(\sum f_{v}\) is the person frequencies or number of persons obtaining each person ability score.

Equation 4 describes the algebraic computations for obtaining the variance \((Uv)\) of the distributions of person logit correct scores \((Sv)\). \((Uv)\) is computed by squaring the mean person logit correct \((X.v^{2})\) and multiplying this value by the person frequency for each possible person ability score \((f_{v})\). The obtained values are summated across the different person ability score groups to produce the value \((\sum f_{v}X.^{2})\). This value \((\sum f_{v}X.^{2})\) is then subtracted from the value \((\sum f_{v}X_{v}^{2})\), which is the sum of the person frequencies \((f_{v})\) times the squared person logit correct values \((X_{v}^{2})\) listed in column 8. This value is then divided by the number of persons included in the edited person-item data matrix minus 1, that is, \((\sum f_{v}-1)\).
\[ Uv = \frac{\sum_{i} f_i v_{Xv}^2 - \sum_{i} f_i v_{Xv} \cdot \sum_{i} f_i v_{Xv} \cdot 2}{\sum_{i} f_i \cdot 2 - 1} \]  

Equation 4

Where:

- \( Uv \) is the variance of the distribution of person logit correct scores,
- \( \sum f_i v_{Xv}^2 \) is the sum of the person frequencies times the squared person logit correct values,
- \( \sum f_i v_{Xv} \cdot 2 \) is the sum of the person frequencies times the square of the mean person logit correct value, and
- \( \sum f_i \cdot 2 - 1 \) is the number of persons included in the edited person-item matrix minus 1.

The equations for arriving at the means and variances of the distributions of person logit correct values along with numerical examples of these computations are presented at the bottom of Table 4. As shown in column 9 of Table 4, the obtained value for the initial person ability measure (\( b v^0 \)) is the same as the obtained person logit correct value (\( Xv \)) listed in column 6. At this time, no adjustments are made to these values.

The procedures presented in Tables 3 and 4 convert values for the proportion incorrect (item \( p \)-values) and proportion correct (person \( p \)-values) to item logit incorrect and person logit correct scores using logarithmic conversions. Wright and Stone (1979, p. 36) present a table of these conversions and for convenience these conversions are reproduced in Table 5.

To correct the initial item difficulty estimates (\( b_i^0 \)) for effects due to sample spread, and the initial person ability estimates (\( b v^0 \)) for effects due to test width, expansion factors are computed. Equations 5 and 6 show the algebraic computations for the person ability (\( X \)) and item difficulty (\( Y \)) expansion factors.
\[ X = \frac{1 + U_i/2.89}{1 - U_i U_v/8.35}^{1/2} \]  \hspace{1cm} \text{Equation 5}

\[ Y = \frac{1 + U_v/2.89}{1 - U_i U_v/8.35}^{1/2} \]  \hspace{1cm} \text{Equation 6}

where:

- \( X \) is the person ability expansion factor due to test width,
- \( Y \) is the item difficulty expansion factor due to sample spread,
- \( U_i \) is the variance of the distribution of item logit incorrect values,
- \( U_v \) is the variance of the distribution of person logit correct scores.

The other values are constants. The computations for \( U_i \) and \( U_v \) were presented in Tables 3 and 4.

Table 6 presents the corrected item difficulties obtained on the 34 persons included in the edited item-person matrix. Column 1 identifies the item score group (\( i \)) and column 2 the names of the items in each score group. Column 3 reports the initial item calibration values (\( b_i^0 \)); these are the same values listed in column 10 of Table 3. The sample spread expansion factor (\( Y \)) is presented in column 4; note that this factor remains constant for all items. In column 5, this factor (\( Y \)) is applied to the initial item calibration values (\( b_i^0 \)) to obtain the corrected item calibration value (\( b_i \)). Column 6 lists for convenience, the item scores associated with each of the corrected item calibration values (\( b_i \)).
Equation 7 describes the algebraic computations used to obtain the standard error of the corrected item difficulty estimates \( SE(b_i) \).

\[
SE(b_i) = Y \left[ \frac{N}{s_i} \left( \frac{N-s_i}{s_i} \right) \right]^{1/2} \quad \text{Equation 7}
\]

where:

- \( SE(b_i) \) is the standard error of measurement for the corrected item difficulty estimate,
- \( Y \) is the item difficulty expansion factor due to sample spread,
- \( N \) is the number of examinees included in the edited item-person matrix, and
- \( s_i \) is the item score or the number of examinees who correctly answered the test item.

The standard error of measurements for the corrected item difficulty estimates, \( SE(b_i) \), are listed in column 7 of Table 6.

Table 7 presents the final estimates of person ability scores for all possible scores on the 14 item test. Column 1 identifies the possible test score \( v \). Again, 13 scores were possible for the 14 item test which comprised the edited person-item matrix. Column 2 presents the initial person measures \( \theta_v \) obtained from column 9 of Table 7. The test width expansion factor \( X \) is identified in column 3; note that this factor also remains constant across all possible person scores. In column 4, this expansion factor \( X \) is applied to the initial person ability measures \( \theta_v \) to obtain a corrected person ability estimate \( \theta_v' \).
Equation 8 describes the algebraic computations used to obtain the standard error of the corrected person ability measure \( SE(bv) \).

\[
SE(bv) = X \left[ \frac{n}{v(n-v)} \right]^{1/2}
\]

Equation 8

where:

- \( SE(bv) \) is the standard error of measurement for the corrected person ability estimate,
- \( X \) is the person ability expansion factor due to test width,
- \( n \) is the number of items included in the edited person-item matrix, and
- \( v \) is the person ability score or the number of items correctly answered by the examinee.

The procedures which have been presented in Appendix A comprise PROX (Cohen, 1976; Wright & Stone, 1979). PROX is an algebraic formulation of the one-parameter model. These procedures provide fairly accurate approximations of the person ability and item difficulty parameter estimates obtained with the more sophisticated maximum likelihood one-parameter models. Also, they embody in a logical and straightforward manner the basic theoretical tenets and mathematical computations of item response models. As such, they provide a readily interpretable framework for introducing this approach to the layperson of psychometric theory. These procedures are less restrictive, in terms of the sample sizes required for item calibration, and as such, they are applicable to data obtained in industrial-organizational studies where typically, only small sample sizes are available.
Table 1
The Unedited Person-Item Matrix

<table>
<thead>
<tr>
<th>ITEM NAME</th>
<th>PERSON NAME</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 2
The Edited and Ordered Person-Item Matrix

<table>
<thead>
<tr>
<th>ITEM NAME</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>EDITED ITEM SCORE</th>
<th>PROPORTION OF 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**EDITED ITEM SCORE**
32 31 31 30 30 27 24 12 7 6 3 1 1 1

**PROPORTION OF 16**
.94 .91 .91 .88 .88 .79 .71 .35 .21 .18 .09 .03 .03 .03
Table 3

Grouped Distribution of the 10 Different Item Scores Obtained on the 34 Persons.

<table>
<thead>
<tr>
<th>ITEM</th>
<th>SCORE</th>
<th>GROUP</th>
<th>ITEM INDEX</th>
<th>NAME</th>
<th>SCORE</th>
<th>FREQUENCY</th>
<th>CORRECT</th>
<th>INCORRECT</th>
<th>INCORRECT SQUARE</th>
<th>X LOGIT</th>
<th>CALIBRATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>32</td>
<td>1</td>
<td>.94</td>
<td>.06</td>
<td>-2.75*</td>
<td>-2.75</td>
<td>7.16</td>
<td>-2.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5,7</td>
<td>31</td>
<td>2</td>
<td>.91</td>
<td>.09</td>
<td>-2.31</td>
<td>-4.62</td>
<td>10.67</td>
<td>-2.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6,9</td>
<td>30</td>
<td>2</td>
<td>.88</td>
<td>.12</td>
<td>-1.99</td>
<td>-3.98</td>
<td>7.92</td>
<td>-2.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>27</td>
<td>3</td>
<td>.79</td>
<td>.21</td>
<td>-1.32</td>
<td>-1.32</td>
<td>1.74</td>
<td>-1.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>24</td>
<td>1</td>
<td>.71</td>
<td>.29</td>
<td>-0.90</td>
<td>-0.90</td>
<td>0.81</td>
<td>-1.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>12</td>
<td>1</td>
<td>.35</td>
<td>.65</td>
<td>+0.62</td>
<td>+0.62</td>
<td>0.38</td>
<td>+0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>7</td>
<td>1</td>
<td>.21</td>
<td>.79</td>
<td>+1.32</td>
<td>+1.32</td>
<td>1.74</td>
<td>+1.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>.18</td>
<td>.82</td>
<td>+1.32</td>
<td>+1.32</td>
<td>2.31</td>
<td>+1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>3</td>
<td>1</td>
<td>.09</td>
<td>.91</td>
<td>+2.31</td>
<td>+2.31</td>
<td>5.34</td>
<td>+2.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15,16,17</td>
<td>1</td>
<td>3</td>
<td>.03</td>
<td>.97</td>
<td>+3.48</td>
<td>+3.48</td>
<td>10.44</td>
<td>+3.29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*These values come from Table 5 where \( \log(0.06/0.94) = -2.75 \).
Table 4

Grouped Distribution of Observed Person Scores on 14 Items

<table>
<thead>
<tr>
<th>PERSON INDEX</th>
<th>PERSON ABILITY SCORE GROUP</th>
<th>PERSON ABILITY PERSON SCORE</th>
<th>POSSIBLE FREQUENCY</th>
<th>PERSON PROPORTION CORRECT</th>
<th>PROPORTION INCORRECT</th>
<th>LOGIT CORRECT</th>
<th>FREQUENCY X LOGIT</th>
<th>INITIAL PROPORTION</th>
<th>PROPORTION LO G I T</th>
<th>FREQUENCY SQUARED X LOGIT</th>
<th>INITIAL LOGIT PERSON MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.07</td>
<td>.93</td>
<td>-2.59*</td>
<td>0.00</td>
<td>0.00</td>
<td>-2.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>.14</td>
<td>.86</td>
<td>-1.82</td>
<td>-1.82</td>
<td>3.31</td>
<td>-1.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>.21</td>
<td>.79</td>
<td>-1.32</td>
<td>-2.66</td>
<td>3.48</td>
<td>-1.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>.29</td>
<td>.71</td>
<td>-0.90</td>
<td>-1.80</td>
<td>1.62</td>
<td>-0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>.36</td>
<td>.64</td>
<td>-0.58</td>
<td>-1.16</td>
<td>0.67</td>
<td>-0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>.43</td>
<td>.57</td>
<td>-0.28</td>
<td>-0.84</td>
<td>0.24</td>
<td>-0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>12</td>
<td>.50</td>
<td>.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>5</td>
<td>.57</td>
<td>.43</td>
<td>0.28</td>
<td>1.40</td>
<td>0.39</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>4</td>
<td>.64</td>
<td>.36</td>
<td>0.58</td>
<td>2.32</td>
<td>1.35</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>1</td>
<td>.71</td>
<td>.29</td>
<td>0.90</td>
<td>0.90</td>
<td>0.81</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>2</td>
<td>.79</td>
<td>.21</td>
<td>1.32</td>
<td>2.54</td>
<td>3.48</td>
<td>1.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>0</td>
<td>.86</td>
<td>.14</td>
<td>1.82</td>
<td>0.00</td>
<td>0.00</td>
<td>1.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0</td>
<td>.93</td>
<td>.07</td>
<td>2.59</td>
<td>0.00</td>
<td>0.00</td>
<td>2.59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ X_v = \frac{f_v}{\Sigma f_v/\sqrt{N}} \]

\[ U_v = \left(\frac{f_v}{\Sigma f_v^2/\sqrt{(\Sigma f_v)^2}} \right) \]

\[ X_v = -1.00/34 = -0.03 \]

\[ U_v = (15.35 - 0.03)/(33) = 0.46 \]

*These values come from Table 2.4.3 where ln[.06/.94] = -2.75.
Table 5  
Proportion to Logit Conversions

Logit = In [Proportion/(1 - Proportion)]

<table>
<thead>
<tr>
<th>PROPORTION 1</th>
<th>LOGIT</th>
<th>PROPORTION</th>
<th>LOGIT</th>
<th>PROPORTION</th>
<th>LOGIT</th>
<th>PROPORTION</th>
<th>LOGIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>-4.60</td>
<td>.26</td>
<td>-1.05</td>
<td>.51</td>
<td>0.04</td>
<td>.76</td>
<td>1.15</td>
</tr>
<tr>
<td>.02</td>
<td>-3.89</td>
<td>.27</td>
<td>-0.99</td>
<td>.52</td>
<td>0.08</td>
<td>.77</td>
<td>1.21</td>
</tr>
<tr>
<td>.03</td>
<td>-3.48</td>
<td>.28</td>
<td>-0.94</td>
<td>.53</td>
<td>0.12</td>
<td>.78</td>
<td>1.27</td>
</tr>
<tr>
<td>.04</td>
<td>-3.18</td>
<td>.29</td>
<td>-0.90</td>
<td>.54</td>
<td>0.16</td>
<td>.79</td>
<td>1.32</td>
</tr>
<tr>
<td>.05</td>
<td>-2.94</td>
<td>.30</td>
<td>-0.85</td>
<td>.55</td>
<td>0.20</td>
<td>.80</td>
<td>1.39</td>
</tr>
<tr>
<td>.06</td>
<td>-2.75</td>
<td>.31</td>
<td>-0.80</td>
<td>.56</td>
<td>0.24</td>
<td>.81</td>
<td>1.45</td>
</tr>
<tr>
<td>.07</td>
<td>-2.59</td>
<td>.32</td>
<td>-0.75</td>
<td>.57</td>
<td>0.28</td>
<td>.82</td>
<td>1.52</td>
</tr>
<tr>
<td>.08</td>
<td>-2.44</td>
<td>.33</td>
<td>-0.71</td>
<td>.58</td>
<td>0.32</td>
<td>.83</td>
<td>1.59</td>
</tr>
<tr>
<td>.09</td>
<td>-2.31</td>
<td>.34</td>
<td>-0.66</td>
<td>.59</td>
<td>0.36</td>
<td>.84</td>
<td>1.66</td>
</tr>
<tr>
<td>.10</td>
<td>-2.20</td>
<td>.35</td>
<td>-0.62</td>
<td>.60</td>
<td>0.41</td>
<td>.85</td>
<td>1.73</td>
</tr>
<tr>
<td>.11</td>
<td>-2.09</td>
<td>.36</td>
<td>-0.58</td>
<td>.61</td>
<td>0.45</td>
<td>.86</td>
<td>1.82</td>
</tr>
<tr>
<td>.12</td>
<td>-1.99</td>
<td>.37</td>
<td>-0.53</td>
<td>.62</td>
<td>0.49</td>
<td>.87</td>
<td>1.90</td>
</tr>
<tr>
<td>.13</td>
<td>-1.90</td>
<td>.38</td>
<td>-0.49</td>
<td>.63</td>
<td>0.53</td>
<td>.88</td>
<td>1.99</td>
</tr>
<tr>
<td>.14</td>
<td>-1.82</td>
<td>.39</td>
<td>-0.45</td>
<td>.64</td>
<td>0.58</td>
<td>.89</td>
<td>2.09</td>
</tr>
<tr>
<td>.15</td>
<td>-1.73</td>
<td>.40</td>
<td>-0.41</td>
<td>.65</td>
<td>0.62</td>
<td>.90</td>
<td>2.20</td>
</tr>
<tr>
<td>.16</td>
<td>-1.66</td>
<td>.41</td>
<td>-0.36</td>
<td>.66</td>
<td>0.66</td>
<td>.91</td>
<td>2.31</td>
</tr>
<tr>
<td>.17</td>
<td>-1.59</td>
<td>.42</td>
<td>-0.32</td>
<td>.67</td>
<td>0.71</td>
<td>.92</td>
<td>2.44</td>
</tr>
<tr>
<td>.18</td>
<td>-1.52</td>
<td>.43</td>
<td>-0.28</td>
<td>.68</td>
<td>0.75</td>
<td>.93</td>
<td>2.59</td>
</tr>
<tr>
<td>.19</td>
<td>-1.45</td>
<td>.44</td>
<td>-0.24</td>
<td>.69</td>
<td>0.80</td>
<td>.94</td>
<td>2.75</td>
</tr>
<tr>
<td>.20</td>
<td>-1.39</td>
<td>.45</td>
<td>-0.20</td>
<td>.70</td>
<td>0.85</td>
<td>.95</td>
<td>2.94</td>
</tr>
<tr>
<td>.21</td>
<td>-1.32</td>
<td>.46</td>
<td>-0.16</td>
<td>.71</td>
<td>0.90</td>
<td>.96</td>
<td>3.18</td>
</tr>
<tr>
<td>.22</td>
<td>-1.27</td>
<td>.47</td>
<td>-0.12</td>
<td>.72</td>
<td>0.94</td>
<td>.97</td>
<td>3.48</td>
</tr>
<tr>
<td>.23</td>
<td>-1.21</td>
<td>.48</td>
<td>-0.08</td>
<td>.73</td>
<td>0.99</td>
<td>.98</td>
<td>3.89</td>
</tr>
<tr>
<td>.24</td>
<td>-1.15</td>
<td>.49</td>
<td>-0.04</td>
<td>.74</td>
<td>1.05</td>
<td>.99</td>
<td>4.60</td>
</tr>
<tr>
<td>.25</td>
<td>-1.10</td>
<td>.50</td>
<td>-0.00</td>
<td>.75</td>
<td>1.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For person scores this "proportion" becomes the number of correct responses $Sv$ divided by the number of test items $n$. Thus the person ability logit is $\ln \left( \frac{Sv}{n-Sv} \right) \ln \left( \frac{Sv}{n-Sv} \right)$.  

For item scores this "proportion" becomes the number of incorrect responses $(N - s_l)$ divided by the sample size $N$. Thus the item difficulty logit is $\ln \left( \frac{N - s_l}{N} \right) \ln \left( \frac{N - s_l}{N} \right)$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 6
Corrected Item Difficulties From 34 Persons

<table>
<thead>
<tr>
<th>ITEM SCORE GROUP</th>
<th>ITEM NAME</th>
<th>INITIAL ITEM CALIBRATION</th>
<th>SAMPLE SPREAD EXPANSION FACTOR</th>
<th>CORRECTED ITEM CALIBRATION</th>
<th>ITEM SCORE</th>
<th>CALIBRATION STANDARD ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>-2.94</td>
<td>1.31</td>
<td>-3.85</td>
<td>32</td>
<td>.95</td>
</tr>
<tr>
<td>2</td>
<td>5,7</td>
<td>-2.50</td>
<td>1.31</td>
<td>-3.28</td>
<td>31</td>
<td>.79</td>
</tr>
<tr>
<td>3</td>
<td>6,9</td>
<td>-2.18</td>
<td>1.31</td>
<td>-2.86</td>
<td>30</td>
<td>.70</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>-1.51</td>
<td>1.31</td>
<td>-1.98</td>
<td>27</td>
<td>.56</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>-1.09</td>
<td>1.31</td>
<td>-1.43</td>
<td>24</td>
<td>.49</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>+0.43</td>
<td>1.31</td>
<td>0.56</td>
<td>12</td>
<td>.47</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>+1.13</td>
<td>1.31</td>
<td>1.48</td>
<td>7</td>
<td>.56</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>+1.33</td>
<td>1.31</td>
<td>1.74</td>
<td>6</td>
<td>.59</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>+2.12</td>
<td>1.31</td>
<td>2.78</td>
<td>3</td>
<td>.79</td>
</tr>
<tr>
<td>10</td>
<td>15,16,17</td>
<td>+3.29</td>
<td>1.31</td>
<td>4.31</td>
<td>1</td>
<td>1.33</td>
</tr>
</tbody>
</table>

N = 34

$SE(b_i) = \sqrt{\frac{N}{n_i(N - n_i)}}^{1/2}$
Table 7

Final Estimates of Person Ability Scores for All Possible Scores on the 14 Item Test

<table>
<thead>
<tr>
<th>POSSIBLE TEST SCORE</th>
<th>INITIAL MEASURE</th>
<th>TEST WIDTH EXPANSION FACTOR</th>
<th>CORRECTED MEASURE</th>
<th>MEASURE STANDARD ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>$b_v$</td>
<td>$X$</td>
<td>$b_vXb_v$</td>
<td>$SE(b_v)$</td>
</tr>
<tr>
<td>1</td>
<td>-2.59</td>
<td>2.09</td>
<td>-5.41</td>
<td>2.17</td>
</tr>
<tr>
<td>2</td>
<td>-1.82</td>
<td>2.09</td>
<td>-3.80</td>
<td>1.60</td>
</tr>
<tr>
<td>3</td>
<td>-1.32</td>
<td>2.09</td>
<td>-2.76</td>
<td>1.36</td>
</tr>
<tr>
<td>4</td>
<td>-0.90</td>
<td>2.09</td>
<td>-1.88</td>
<td>1.24</td>
</tr>
<tr>
<td>5</td>
<td>-0.58</td>
<td>2.09</td>
<td>-1.21</td>
<td>1.17</td>
</tr>
<tr>
<td>6</td>
<td>-0.28</td>
<td>2.09</td>
<td>-0.59</td>
<td>1.13</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>2.09</td>
<td>0.00</td>
<td>1.12</td>
</tr>
<tr>
<td>8</td>
<td>0.28</td>
<td>2.09</td>
<td>0.59</td>
<td>1.13</td>
</tr>
<tr>
<td>9</td>
<td>0.90</td>
<td>2.09</td>
<td>1.21</td>
<td>1.17</td>
</tr>
<tr>
<td>10</td>
<td>0.90</td>
<td>2.09</td>
<td>1.88</td>
<td>1.24</td>
</tr>
<tr>
<td>11</td>
<td>1.32</td>
<td>2.09</td>
<td>2.76</td>
<td>1.36</td>
</tr>
<tr>
<td>12</td>
<td>1.82</td>
<td>2.09</td>
<td>3.80</td>
<td>1.60</td>
</tr>
<tr>
<td>13</td>
<td>2.59</td>
<td>2.09</td>
<td>5.41</td>
<td>2.17</td>
</tr>
</tbody>
</table>

$n=14$

$SE(b_v) = X[(n/v)(n-v)]^{1/2}$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX B

The $a_i$, $b_i$ and $c_i$ Item Characteristic Curve Parameter Estimates Obtained by LOGIST 5 on the 18 Data Sets.
Autobiography

James B. Flynn was born in Trenton, NJ on March 24, 1953. He obtained his B. S. in Psychology from Lynchburg College, Lynchburg, VA in 1975. At Lynchburg College he was on the Dean's List all four years and was a member of Phi Kappa Phi (Scholastic Honorary). He graduated from Lynchburg College Cum Laude and with Honors in Psychology. He graduated from Old Dominion University in 1979 with an M. S. in General-Experimental Psychology with an emphasis on Human Factors Psychology. His Masters thesis was entitled "The Effects of Alcohol on the Rate of Gain of Information." While obtaining his doctorate in industrial-organizational psychology, James Flynn acquired five years applied research experience in business and industry, working for such organizations as: Philip Morris, U.S.A.; Calspan Corporation; Ball Foundation; Air Force Human Resource Laboratory; Organization Research Group of Tidewater, Inc.; and ODU's Performance Assessment Laboratory. He is the author of six technical reports and five professional paper presentations. In addition to these experiences, James Flynn has taught undergraduate psychology courses at Old Dominion University. He is a member of the American Psychological Association, Academy of Management, Society of Sigma Xi, and the American Society of Personnel Administrators. Most recently James Flynn was a USAF-SCEEE Graduate Student Summer Fellow.