Comparison of Two Logistic Multidimensional Item Response Theory Models

Research Report ONR90-8

Judith A. Spray, Tim C. Davey, Mark D. Reckase
Terry A. Ackerman
James E. Carlson


Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

October 1990
Comparison of Two Logistic Multidimensional Item Response Theory Models

Judith A. Spray, Tim C. Davey, and Mark D. Reckase
American College Testing

Terry A. Ackerman
University of Illinois

James E. Carlson
Auburn University at Montgomery

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Comparison of two logistic multidimensional item response theory models

Test data generated according to two different multidimensional item response theory models were compared at both the item response level and the test score level to determine if measurable differences between the models could be detected when the data sets were constrained to be equivalent in terms of item p-values. Although differences could be detected at the item level, these differences decreased as the correlation between examinee abilities increased. Furthermore, these item differences were small in magnitude and could not be considered unimportant or insignificant from a practical standpoint. No differences were found at the total test score level, and it was concluded that, at least for the data used in this study, the models were indistinguishable.
Abstract

Test data generated according to two different multidimensional item response theory models were compared at both the item response level and the test score level to determine if measurable differences between the models could be detected when the data sets were constrained to be equivalent in terms of item \( p \)-values. Although differences could be detected at the item level, these differences decreased as the correlation between examinee abilities increased. Furthermore, these item differences were small in magnitude and could be considered unimportant or insignificant from a practical standpoint. No differences were found at the total test score level, and it was concluded that, at least for the data used in this study, the models were indistinguishable.
Comparison of Two Logistic Multidimensional Item Response Theory Models

Psychometricians who have some interest in multidimensional item response theory (MIRT) modeling may be familiar with the terms, *compensatory* and *noncompensatory* as they relate to two general model classification schemes. Ansley and Forsyth (1985) contrasted the two types of model classifications as follows. "Compensatory models, unlike noncompensatory models, permit high ability on one dimension to compensate for low ability on another dimension in terms of probability of correct response. In the noncompensatory models, the minimum factor (probability) in the denominator is the upper bound for the probability of a correct response. Thus, for a two-dimensional item, a person with a very low ability on one dimension and very high ability on the other has a very low probability of correctly answering the item" (p. 40).

Typically, MIRT models of the compensatory type, such as the logistic MIRT model (Doody-Bogan & Yen, 1983; Hattie, 1981; Reckase, 1985, 1986) or the normal ogive MIRT model (Samejima, 1974) imply linear combinations of the multidimensional abilities in the exponent of the expression for the probability of a correct response. In this linear fashion, a low ability on one or more of the $k$ ability dimensions can be compensated by a higher ability on one or more of the remaining dimensions. Because the compensation is a characteristic of this linear combination, such models are probably more accurately labeled *linear MIRT* models. A typical linear logistic MIRT model of the compensatory type can be written as

$$P_j(\Theta) = c_j + (1 - c_j) \frac{\sum_{i} f_{ijm} \cdot d_j}{1 + e^{\sum_{i} f_{ijm} \cdot d_j}}.$$  \hspace{1cm} (1)
where

\[ f_{ijm} = a_{jm} \theta_{im} \]

- \( f_{ijm} \) = the pseudo-guessing parameter of the \( j \)th item,
- \( a_{jm} \) = the discrimination parameter for the \( j \)th item on the \( m \)th dimension,
- \( d_{j} \) = the difficulty parameter for the \( j \)th item, and
- \( \theta_{im} \) = the \( m \)th element in the \( i \)th person's ability vector, \( \theta_i \).

In this model the favorable response probability, \( P_j(\theta_i) \), is bounded from below by \( c_j \). However, because the upper bound of \( P_j(\theta_i) \) is not a function of any one ability dimension, it increases monotonically as \( \sum_{m=1}^{k} f_{ijm} \) increases.

On the other hand, noncompensatory MIRT models (Sympson, 1978; Embretson, 1984) describe the probability of a favorable response in terms of a product of \( k \) functions of ability on a single dimension and item characteristics. In its most common form, a logistic MIRT model of this noncompensatory or multiplicative type can be written as

\[
P_j(\theta_i) = c_j + (1-c_j) \prod_{m=1}^{k} \frac{e^{f_{ijm}}}{1+e^{f_{ijm}}},
\]

where now we let \( f_{ijm} = [a_{jm} (\theta_{im} - b_{jm})] \) with \( b_{jm} \) = the difficulty parameter for the \( j \)th item on the \( m \)th dimension. \( P_j(\theta_i) \) is bounded by an upper asymptote equal to the minimum of \( \exp\{f_{ijm}\}/(1+\exp\{f_{ijm}\}) \), and the lower asymptote, \( c_j \), for any given examinee with \( \theta = \theta_i \). Thus, the noncompensatory nature of the model is due to the fact that \( P_j(\theta_i) \) can never be greater than the minimum value of the terms in the product, \( \exp\{f_{ijm}\}/(1+\exp\{f_{ijm}\}) \), a function of the smallest value of the \( k \)
ability dimensions for a given examinee. Because of its multiplicative form, the model is more generally labeled as a multiplicative MIRT model.

Researchers have used the multiplicative MIRT model to examine characteristics of unidimensional item response theory parameter estimates derived from MIRT response data (Ansley & Forsyth, 1985) and to model certain multicomponent latent traits in response processes (Embretson, 1984). Reckase (1985) has used a linear MIRT model on real response data to estimate two-dimensional item and person parameters on an ACT Assessment Mathematics Usage test. However, no one has actually shown that one model is more representative of the actual item-examinee response process than the other. It may even be possible that one model may be appropriate under one set of circumstances while the other type may be more appropriate in other situations.

In this paper we investigate the differences between item responses generated by these two logistic MIRT models. We have been interested in determining whether or not it is possible to distinguish one model or process from the other through some evaluation of response data. More specifically, our concern has been in establishing whether or not it is possible to detect differences between these two MIRT models, either at the item response or test score level, when the item parameters from each MIRT model have been matched or equated in some sense.

The first task was to establish the item parameters from one of the logistic MIRT models that would produce "reasonable" $p$-values or proportion-correct indices for a specified examinee population. Therefore, a target distribution of $p$-values for a 20-item test was conceived and item parameters for a linear or compensatory MIRT model were chosen, basically by trial-and-error, until the expected $p$-value with respect to this examinee population matched the target distribution. Table 1 gives the set of item parameters for the 20 items for the model given by equation (1). The table also gives the expected value of each $p$-value under the assumption that the ability vector, $\theta$, for the examinee population,
was distributed as bivariate normal with mean vector, \( \mathbf{0} \), and variance-covariance matrix of ones along the diagonal and with nondiagonal values equal to \( \rho \) (.00, .25, .50, or .75). All c-parameters were set to zero.

In order to produce a comparable or "matched" set of noncompensatory, or multiplicative model item parameters, estimates of these item parameters were obtained by minimizing

\[
\sum_{i=1}^{N} \left[ P_C(\mathbf{\theta}, \mathbf{a}, \mathbf{d}) - P_{NC}(\mathbf{\theta}, \mathbf{\hat{a}}, \mathbf{\hat{b}}) \right]^2
\]

for \( N = 2000 \) randomly selected examinees with ability, \( \mathbf{\theta} \), distributed as given previously, where \( P_C \) and \( P_{NC} \) represent logistic MIRT models given by equations (1) and (2), respectively. This process was repeated for 10 replications for each of \( k = 1, 2, \ldots, 20 \) items to ensure that the estimates obtained weren't unduly influenced by the samples selected or the starting values used. Mean values of the replication estimates yielded the noncompensatory item parameters listed in Tables 2-5, for \( \rho \) values of .00, .25, .50, and .75. The expected value of each item's \( p \)-value is given in the last column of each table. Because the least squares minimization procedure produces unbiased estimates of \( P_{NC} \), the expected value of each \( p \)-value under the noncompensatory model should be equal to that of the compensatory model, within some estimation error. Equivalence of \( p \)-values was the critical matching criterion between the two MIRT models.


Model Differences at the Total Test Level

By treating the two sets of item parameters as known for each of the two MIRT models, we first investigated the differences between expected number-correct score frequencies of a 20-item test when $\theta$ was distributed as a bivariate normal random vector with distributions given previously. These frequencies were estimated by evaluating either the number-correct distribution under the compensatory model, $h_C(y)$ or the noncompensatory model, $h_{NC}(y)$, for $y = 0, 1, 2, \ldots, 20$, or

$$h_C(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_c(y \mid \theta) \ g(\theta) \ d\theta_1 \ d\theta_2 \quad (4)$$

and

$$h_{NC}(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{NC}(y \mid \theta) \ g(\theta) \ d\theta_1 \ d\theta_2 . \quad (5)$$

In each case, the conditional frequencies, $f_c(y \mid \theta)$ and $f_{NC}(y \mid \theta)$, were computed using either models (1) or (2), and a recursive procedure described by Lord and Wingersky (1984). Table 6 gives the signed differences between the frequencies, $h_C(y) - h_{NC}(y)$, for $y = 0, 1, 2, \ldots, 20$, for rho values of .00, .25, .50, and .75. The greatest differences, as expected, occurred for the highest number-correct scores, but the differences in frequencies were small, never greater than .015. For most number-correct score values, these differences became smaller as rho increased.
Another way to assess the significance of these differences was to determine how much data would need to be observed before the differences were statistically detectable. This was done by calculating the minimum sample size required to reject the homogeneity of parallel populations with given levels of test significance and power. These calculations assumed a multivariate normal approximation for each model's multinomial distribution of observed-score frequencies which in turn produced the quadratic form of the noncentrality parameter of a noncentral chi square distribution. The minimum sample size followed as a direct function of this parameter, the specified test significance, and power. For example, with a significance level of .01 and power equal to .95, the minimum sample sizes were 1678, 3242, 7466, and 15311 for correlated ability distributions with rho equal to .00, .25, .50, and .75, respectively. These sample sizes state that even in the unlikely event of uncorrelated ability distributions, it would still require at least 1678 observed scores from both the compensatory and noncompensatory MIRT models before the null hypothesis of model equivalence could be rejected with a power of .95.

The first four (central) moments of each number-correct distribution are given in Table 7 for each value of rho. Both distributions were negatively skewed with the compensatory distribution slightly more platykurtic and both were generally flatter than the normal distribution. The variances of the number-correct scores increased with an increase in rho, and in general, the distributions of number-correct scores became increasingly similar as rho increased.
A contour plot of the (signed) difference between the number-correct true scores under the two models, or

\[ \sum_{j=1}^{20} P_{jc}(\theta) - \sum_{j=1}^{20} P_{jnc}(\theta) \]

was another way to observe model differences at the total test level for various \((\theta_1, \theta_2)\) points in the ability space. The greatest differences occurred when either \(\theta_1\) or \(\theta_2\) was low. See Figures 1-4 for rho values of .00, .25, .50, and .75, respectively. It should be noted that, in these plots, the only influence of rho was through the values of the noncompensatory item parameters. Recall that the compensatory item parameters were fixed for all values of rho. Therefore, when interpreting these contour plots, one has to mentally superimpose the appropriate bivariate normal distribution over the contours in order to evaluate the importance of the true-score differences observed.

---

Insert Figures 1-4 Here

---

Another way to compare the two MIRT models was to observe the amount of multidimensional information (MINF) for different points in the ability space between the two models. MINF has been defined (Reckase, 1986) as a direct generalization of the unidimensional IRT concept of item information (i.e., the ratio of the square of the slope of the item characteristic curve at an ability point, \(\theta\), to the variance of the error of the item score at that level of \(\theta\)). For the definition of MINF, the slope of the item characteristic surface must be evaluated in a particular direction, \(\alpha\), a vector of angles with the coordinate axes of the ability space.

Plots of the absolute difference between the compensatory and noncompensatory test information vectors (i.e, the sum of item information across
the 20 items) for item parameters estimated with rho values of .00, .25, .50, and .75 (Figures 5-8, respectively) showed that model differences might be significant if abilities were negatively correlated. However, for all "likely" ability distributions, there were no meaningful differences in MINF between the two models, and these absolute differences appeared to decrease as rho increased.

Model Differences at the Item Level

It was also of interest to evaluate the differences between models at the single item response level. There were two ways in which this was done. The first involved the evaluation of the ideal observer index (Davey, Levine, & Williams, 1989; Levine, Drasgow, Williams, McCusker, & Thomasson, 1990). A more complete definition of this index is provided in the appendix of this paper. However, a simplified definition is as follows. The ideal observer index (IOI) is a measure of the proportional number of times that a correct decision is made concerning which of the two competing models produced a particular response to an item. The decision is one that is made hypothetically by an "ideal observer," or an individual who has access to all of the information necessary to yield the highest possible percent of model classification (i.e., compensatory vs. noncompensatory). As far as the ideal observer is concerned, if the item response data fail to distinguish between the two competing models, then the value of this index would be at or near the chance level of .5. Conversely, readily distinguishable models should yield an index near 1.0.

Table 8 shows that the IOI was greater than chance, implying that there was a difference between the models for all 20 items. However, the IOI was never greater than .60 and was greater than .55 for only three items, numbers 3, 6,
and 7, when rho was .00. The value of the IOI decreased for each item as rho increased, implying that it became more difficult to distinguish between the models as the correlation coefficient increased.

One way to think of the magnitude of the IOI was to imagine how many trials of the IO experiment would be necessary before the ideal observer could ascertain, with some given level of certainty, that the models were actually distinguishable. This would be comparable to a test of the difference between any obtained IOI from Table 8 and the null proportion of correct model classifications due to chance. For example, to be able to detect a true difference between the models for item number 6 with a zero value of rho would require at least 40 trials of the IO experiment. This would be comparable to a test of the null proportion of correct classifications due to chance or .50 versus the (true) alternative proportion (.555) with a significance of .01 and power of .95. Conversely, a true IOI of .52 would require more than 290 trials at similar levels of test significance and power.

-------------------
Insert Table 8 Here
-------------------

Another way to evaluate model differences at the item level was to use a generalized MIRT model, or a reparameterization of both the compensatory and noncompensatory models into a single MIRT model, or

$$P_j(\theta_i) = c_j + (1-c_j) \frac{e^{\ell_{ij1}^* + \ell_{ij2}^*}}{1 + e^{\ell_{ij1}^* + \ell_{ij2}^*} + \mu e^{\ell_{ij1}^* + e^{\ell_{ij2}^*}},}$$

(6)

where $\mu$ represented an indicator variable such that

$$\mu = \begin{cases} 
0, & \text{for the linear or compensatory MIRT model,} \\
1, & \text{for the multiplicative or noncompensatory MIRT model.}
\end{cases}$$
Item response data, $x_{ij}$, were generated from samples of size $N = 2000$ of $\theta_i$ drawn from the bivariate normal distributions mentioned previously. The response data were known to have been produced by either the compensatory or noncompensatory MIRT model and were simulated by comparing the known values of $P_j(\theta_i)$ to a pseudorandomly drawn uniform deviate, $\omega$, such that

$$x_{ij} = \begin{cases} 
1, & 0 \leq \omega < P_j(\theta_i) \\
0, & P_j(\theta_i) \leq \omega < 1.
\end{cases}$$

The least squares estimation procedure was used to estimate the generalized MIRT model parameters. Each estimation was replicated 10 times with randomly selected starting values. Either four or five unique item parameters were estimated from the generalized MIRT model, as given by equation (6). The same item parameters that were given in tables 1-5 were used to generate the response data for the estimation procedure. When the response data were generated by the compensatory model, $a_1, a_2, b$, and $d$ (i.e., $d = -a_1b_1 - a_2b_2$) as well as $\mu$, were estimated. When the response data were generated by the noncompensatory model, $a_1, a_2, b_1, b_2$, and $\mu$ were estimated.

Table 9 shows the average bias in the item parameter estimates and the standard deviations of the estimates (in parentheses). For compensatory data, the model parameter, $\mu$ was estimated fairly accurately for the uncorrelated situation, but the amount of bias and the standard deviation of the estimates increased as rho increased. A similar situation occurred with noncompensatory data. However, although the amount of estimation error increased as the correlation between the abilities increased, the model still remained identifiable, in the sense that for compensatory data, the $\mu$ estimates were statistically "close" to zero.
Likewise, for noncompensatory data, the $\mu$ estimates were statistically "close" to one.

The IOI analysis and the generalized MIRT model estimation gave similar results. That is, there were model differences at the item level, but these differences tended to decrease as the correlation in abilities increased. The generalized MIRT analysis also suggested that these differences might still be estimable, however, even when abilities are strongly correlated.

Summary and Conclusions

These analyses and results seem to indicate that even though it is difficult to observe model differences at the overall test score level, there still may be measurable differences between the responses at the item level. Because the matching criterion between the two models resulted in similar expected $p$-values, we anticipated small differences at the total test score response level, or at the true score level. The differences that were detected at this level were consistent with the differences implied in the two models. Fewer high, number-correct scores or estimated true scores were observed from the noncompensatory model, but these and other total test differences decreased as rho increased. As for the item response level analysis, both the IOI and the generalized MIRT model estimation showed that it is possible to quantify these differences and to distinguish between the data generated by carefully matched item response models of these two types. However, these differences, although real, are very small and probably not significant from any practical standpoint.
Although it is difficult to generalize beyond the two-dimensional situation used in the present study, it would appear to be difficult to distinguish between the two models without the benefit of any prior knowledge of item parameters or abilities. Even with such prior knowledge, response data generated by the models are nearly indistinguishable, especially with correlated abilities, which is likely the case in many real testing situations.
References


Appendix

Analytical Definition of the Ideal Observer Index

A hypothetical observer is presented with two abilities, \( t_1 \) and \( t_2 \), each with their associated item responses, \( u_1 \) and \( u_2 \). The observer is informed that one ability-response pair was generated by one of two competing item response models, while the other pair was generated under the second model. The task is to correctly match each ability-response pair with the proper generating model. To make this decision, the observer is given access to both competing item response functions, \( P_1 \) and \( P_2 \), and the common ability distribution, \( f(t) \).

An ideal observer bases this decision on an optimal rule, \( \delta \), which is determined by the ratio of likelihood functions, \( L_i(t_j,u_j) = P_i(t_j)^{u_j} Q_i(t_j)^{1-u_j} \), where \( Q_i(t_j) = 1 - P_i(t_j), \) \( i = 1, 2; \) \( j = 1, 2 \). The decision rule, \( \delta \), is then defined as

\[
\delta = \begin{cases} 
\text{if } L_1(t_1,u_1) \cdot L_2(t_2,u_2) > L_1(t_2,u_2) \cdot L_2(t_1,u_1), & \text{then decide model } \{P_1;f\} \text{ produced sample } \{t_1,u_1\} \text{ while model } \{P_2;f\} \text{ produced } \{t_2,u_2\}. \\
\end{cases}
\]

The probability of this decision rule being correct, given the model, is

\[
\text{Prob}[\delta \text{ correct} | \text{model}] = \text{Prob}[L_1(t_1,u_1) \cdot L_2(t_2,u_2) > L_1(t_2,u_2) \cdot L_2(t_1,u_1) | \{P_1;f\}&\{P_2;f\}] + \text{Prob}[L_1(t_2,u_2) \cdot L_2(t_1,u_1) > L_1(t_1,u_1) \cdot L_2(t_2,u_2) | \{P_2;f\}&\{P_1;f\}].
\]
The response pair, \( u = (u_1, u_2) \), can be defined in four possible patterns: (1,1), (1,0), (0,1), and (0,0). Therefore,

\[
\text{Prob}\left[ L_1(t_1, u_1) \cdot L_2(t_2, u_2) > L_1(t_2, u_2) \cdot L_2(t_1, u_1) \right| \{P_1; f\} \cup \{P_2; f\} ] = \\
\text{Prob}\left[ P_1(t_1) \cdot P_2(t_2) > P_1(t_2) \cdot P_2(t_1) \right| u = (1,1) ] \cdot \text{Prob}[u = (1,1)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}\left[ P_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot P_2(t_1) \right| u = (1,0) ] \cdot \text{Prob}[u = (1,0)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}\left[ Q_1(t_1) \cdot P_2(t_2) > Q_1(t_2) \cdot Q_2(t_1) \right| u = (0,1) ] \cdot \text{Prob}[u = (0,1)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}\left[ Q_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot Q_2(t_1) \right| u = (0,0) ] \cdot \text{Prob}[u = (0,0)] \cup \{P_1; f\} \cup \{P_2; f\} ].
\]

Define \( \pi_{ij} = \int P_1(t)^{u_1}Q_1(t)^{1-u_1}P_2(g)^{u_2}Q_2(g)^{1-u_2} f(t) f(g) \, dt \, dg \).

Then, \( \text{Prob}\left[ L_1(t_1, u_1) \cdot L_2(t_2, u_2) > L_1(t_2, u_2) \cdot L_2(t_1, u_1) \right| \{P_1; f\} \cup \{P_2; f\} ] = \\
\pi_{11} \cdot \text{Prob}[P_1(t_1) \cdot P_2(t_2) > P_1(t_2) \cdot P_2(t_1)] \right| u = (1,1) ] + \\
\pi_{10} \cdot \text{Prob}[P_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot P_2(t_1)] \right| u = (1,0) ] + \\
\pi_{01} \cdot \text{Prob}[Q_1(t_1) \cdot P_2(t_2) > P_1(t_2) \cdot Q_2(t_1)] \right| u = (0,1) ] + \\
\pi_{00} \cdot \text{Prob}[Q_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot Q_2(t_1)] \right| u = (0,0) ].
\]

Similarly, \( \text{Prob}\left[ L_1(t_2, u_2) \cdot L_2(t_1, u_1) > L_1(t_1, u_1) \cdot L_2(t_2, u_2) \right| \{P_2; f\} \cup \{P_1; f\} ] = \\
\text{Prob}[P_1(t_2) \cdot P_2(t_1) > P_1(t_1) \cdot P_2(t_2) \right| u = (1,1) ] \cdot \text{Prob}[u = (1,1)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}[P_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot P_2(t_2) \right| u = (1,0) ] \cdot \text{Prob}[u = (1,0)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}[Q_1(t_2) \cdot P_2(t_1) > P_1(t_1) \cdot Q_2(t_2) \right| u = (0,1) ] \cdot \text{Prob}[u = (0,1)] \cup \{P_1; f\} \cup \{P_2; f\} ] \\
+ \text{Prob}[Q_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot Q_2(t_2) \right| u = (0,0) ] \cdot \text{Prob}[u = (0,0)] \cup \{P_1; f\} \cup \{P_2; f\} ].
Then, \( \text{Prob}[L_1(t_2, u_2) \cdot L_2(t_1, u_1) > L_1(t_1, u_1) \cdot L_2(t_2, u_2) | \{P_2; f\} \& \{P_1; f\}] = \)

\[
\begin{align*}
\pi_{11} \text{ Prob}[P_1(t_2) \cdot P_2(t_1) > P_1(t_1) \cdot P_2(t_2) | u = (1, 1)] + \\
\pi_{10} \text{ Prob}[P_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot P_2(t_2) | u = (1, 0)] + \\
\pi_{01} \text{ Prob}[Q_1(t_2) \cdot P_2(t_1) > P_1(t_1) \cdot Q_2(t_2) | u = (0, 1)] + \\
\pi_{00} \text{ Prob}[Q_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot Q_2(t_2) | u = (0, 0)].
\end{align*}
\]

Let \( \Omega_{u_1 u_2} \) be defined as that region of the ability space where

\[
P_1(t_1)^{u_1} \cdot Q_1(t_1)^{u_1-1} \cdot P_2(t_2)^{u_2} \cdot Q_2(t_2)^{u_2-1} > P_1(t_1)^{u_2} \cdot Q_1(t_1)^{u_2-1} \cdot P_2(t_2)^{u_1} \cdot Q_2(t_2)^{u_1-1}
\]

holds, and likewise let \( \widetilde{T}_{u_1 u_2} \) be defined as that region of the ability space where

\[
P_1(t_2)^{u_2} \cdot Q_1(t_2)^{u_2-1} \cdot P_2(t_1)^{u_1} \cdot Q_2(t_1)^{u_1-1} > P_1(t_1)^{u_1} \cdot Q_1(t_1)^{u_1-1} \cdot P_2(t_2)^{u_2} \cdot Q_2(t_2)^{u_2-1}
\]

is true. Then

\[
\text{Prob}[P_1(t_1) \cdot P_2(t_2) > P_1(t_2) \cdot P_2(t_1) | u = (1, 1)] = \int\int_{a_{11}} f(t) f(g) \, dt \, dg,
\]

\[
\text{Prob}[P_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot P_2(t_1) | u = (1, 0)] = \int\int_{a_{10}} f(t) f(g) \, dt \, dg,
\]

\[
\text{Prob}[Q_1(t_1) \cdot P_2(t_2) > P_1(t_2) \cdot Q_2(t_1) | u = (0, 1)] = \int\int_{a_{01}} f(t) f(g) \, dt \, dg,
\]

and

\[
\text{Prob}[Q_1(t_1) \cdot Q_2(t_2) > Q_1(t_2) \cdot Q_2(t_1) | u = (0, 0)] = \int\int_{a_{00}} f(t) f(g) \, dt \, dg.
\]
Then

\[ \text{Prob}[P_1(t_2) \cdot P_2(t_1) > P_1(t_1) \cdot P_2(t_2) \mid u = (1,1)] = \int \int f(t) f(g) \, dt \, dg, \]

\[ \text{Prob}[P_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot P_2(t_2) \mid u = (0,1)] = \int \int f(t) f(g) \, dt \, dg, \]

\[ \text{Prob}[Q_1(t_2) \cdot P_2(t_1) > Q_2(t_2) \cdot P_1(t_1) \mid u = (1,0)] = \int \int f(t) f(g) \, dt \, dg, \]

and

\[ \text{Prob}[Q_1(t_2) \cdot Q_2(t_1) > Q_1(t_1) \cdot Q_2(t_2) \mid u = (0,0)] = \int \int f(t) f(g) \, dt \, dg. \]

Thus, \( \text{Prob}[\delta \text{ correct} \mid \text{model}] = \)

\[ \pi_{11} \int \int f(t) f(g) \, dt \, dg + \pi_{10} \int \int f(t) f(g) \, dt \, dg + \]

\[ \pi_{01} \int \int f(t) f(g) \, dt \, dg + \pi_{00} \int \int f(t) f(g) \, dt \, dg + \]

\[ \pi_{11} \int \int f(t) f(g) \, dt \, dg + \pi_{10} \int \int f(t) f(g) \, dt \, dg + \]

\[ \pi_{01} \int \int f(t) f(g) \, dt \, dg + \pi_{00} \int \int f(t) f(g) \, dt \, dg + \]

\[ \pi_{11} + \pi_{10} \int \int f(t) f(g) \, dt \, dg + \int \int f(t) f(g) \, dt \, dg \]

or

\[ \pi_{01} \int \int f(t) f(g) \, dt \, dg + \int \int f(t) f(g) \, dt \, dg \]

\[ \pi_{01} \int \int f(t) f(g) \, dt \, dg + \int \int f(t) f(g) \, dt \, dg + \pi_{00} \]
Finally, \( \text{Prob}[\delta \text{ correct}] = \text{Prob}[\delta \text{ correct}|\text{model}] \cdot \text{Prob[selecting a model]} \). Because each model is equally likely, the probability of selecting a model is equal to .5. Thus, \( \text{Prob}[\delta \text{ correct}] = .5(\text{Prob}[\delta \text{ correct}|\text{model}]) \).
Table 1

*Original Item Parameters for the Compensatory Model*

<table>
<thead>
<tr>
<th>Item #</th>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(d)</th>
<th>(E(p-value))</th>
<th>(\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.25</td>
</tr>
<tr>
<td>01</td>
<td>0.90</td>
<td>1.31</td>
<td>-0.67</td>
<td>.39</td>
<td>.39</td>
</tr>
<tr>
<td>02</td>
<td>2.10</td>
<td>0.50</td>
<td>-1.13</td>
<td>.34</td>
<td>.35</td>
</tr>
<tr>
<td>03</td>
<td>0.89</td>
<td>1.10</td>
<td>0.52</td>
<td>.59</td>
<td>.59</td>
</tr>
<tr>
<td>04</td>
<td>0.99</td>
<td>1.00</td>
<td>-0.44</td>
<td>.42</td>
<td>.42</td>
</tr>
<tr>
<td>05</td>
<td>0.58</td>
<td>1.65</td>
<td>0.78</td>
<td>.63</td>
<td>.62</td>
</tr>
<tr>
<td>06</td>
<td>0.91</td>
<td>1.27</td>
<td>0.42</td>
<td>.57</td>
<td>.57</td>
</tr>
<tr>
<td>07</td>
<td>1.03</td>
<td>0.95</td>
<td>1.08</td>
<td>.69</td>
<td>.68</td>
</tr>
<tr>
<td>08</td>
<td>0.32</td>
<td>2.27</td>
<td>0.38</td>
<td>.55</td>
<td>.55</td>
</tr>
<tr>
<td>09</td>
<td>0.61</td>
<td>0.72</td>
<td>1.63</td>
<td>.80</td>
<td>.79</td>
</tr>
<tr>
<td>10</td>
<td>0.67</td>
<td>1.12</td>
<td>0.60</td>
<td>.51</td>
<td>.61</td>
</tr>
<tr>
<td>11</td>
<td>0.91</td>
<td>0.91</td>
<td>-0.21</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td>12</td>
<td>0.64</td>
<td>1.72</td>
<td>-0.05</td>
<td>.49</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>1.65</td>
<td>0.38</td>
<td>0.40</td>
<td>.57</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>0.18</td>
<td>1.61</td>
<td>1.84</td>
<td>.78</td>
<td>.78</td>
</tr>
<tr>
<td>15</td>
<td>0.82</td>
<td>1.02</td>
<td>0.09</td>
<td>.52</td>
<td>.52</td>
</tr>
<tr>
<td>16</td>
<td>1.45</td>
<td>0.81</td>
<td>-0.24</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td>17</td>
<td>1.64</td>
<td>0.62</td>
<td>0.85</td>
<td>.56</td>
<td>.63</td>
</tr>
<tr>
<td>18</td>
<td>0.77</td>
<td>0.76</td>
<td>-0.91</td>
<td>.32</td>
<td>.33</td>
</tr>
<tr>
<td>19</td>
<td>1.46</td>
<td>0.62</td>
<td>0.10</td>
<td>.52</td>
<td>.52</td>
</tr>
<tr>
<td>20</td>
<td>0.39</td>
<td>1.37</td>
<td>0.32</td>
<td>.56</td>
<td>.56</td>
</tr>
</tbody>
</table>
Table 2

*Item Parameters for the Noncompensatory Model with Rho = .00*

<table>
<thead>
<tr>
<th>Item #</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>E(p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1.26</td>
<td>1.60</td>
<td>-0.92</td>
<td>-0.15</td>
<td>.38</td>
</tr>
<tr>
<td>02</td>
<td>2.30</td>
<td>1.04</td>
<td>0.38</td>
<td>-2.28</td>
<td>.34</td>
</tr>
<tr>
<td>03</td>
<td>1.22</td>
<td>1.39</td>
<td>-1.42</td>
<td>-0.99</td>
<td>.59</td>
</tr>
<tr>
<td>04</td>
<td>1.32</td>
<td>1.35</td>
<td>-0.62</td>
<td>-0.58</td>
<td>.42</td>
</tr>
<tr>
<td>05</td>
<td>1.02</td>
<td>1.82</td>
<td>-2.71</td>
<td>-0.62</td>
<td>.62</td>
</tr>
<tr>
<td>06</td>
<td>1.25</td>
<td>1.53</td>
<td>-1.45</td>
<td>-0.79</td>
<td>.56</td>
</tr>
<tr>
<td>07</td>
<td>1.30</td>
<td>1.26</td>
<td>-1.48</td>
<td>-1.63</td>
<td>.68</td>
</tr>
<tr>
<td>08</td>
<td>0.92</td>
<td>2.38</td>
<td>-3.95</td>
<td>-0.22</td>
<td>.55</td>
</tr>
<tr>
<td>09</td>
<td>0.93</td>
<td>1.00</td>
<td>-2.75</td>
<td>-2.35</td>
<td>.80</td>
</tr>
<tr>
<td>10</td>
<td>1.05</td>
<td>1.37</td>
<td>-1.96</td>
<td>-0.90</td>
<td>.61</td>
</tr>
<tr>
<td>11</td>
<td>1.24</td>
<td>1.25</td>
<td>-0.78</td>
<td>-0.75</td>
<td>.46</td>
</tr>
<tr>
<td>12</td>
<td>1.07</td>
<td>1.92</td>
<td>-2.17</td>
<td>-0.19</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>1.81</td>
<td>0.88</td>
<td>-0.36</td>
<td>-3.25</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>0.85</td>
<td>1.67</td>
<td>-5.26</td>
<td>-1.17</td>
<td>.73</td>
</tr>
<tr>
<td>15</td>
<td>1.17</td>
<td>1.32</td>
<td>-1.21</td>
<td>-0.75</td>
<td>.51</td>
</tr>
<tr>
<td>16</td>
<td>1.71</td>
<td>1.23</td>
<td>-0.27</td>
<td>-1.35</td>
<td>.45</td>
</tr>
<tr>
<td>17</td>
<td>1.83</td>
<td>1.06</td>
<td>-0.68</td>
<td>-2.55</td>
<td>.63</td>
</tr>
<tr>
<td>18</td>
<td>1.09</td>
<td>1.09</td>
<td>-0.31</td>
<td>-0.32</td>
<td>.32</td>
</tr>
<tr>
<td>19</td>
<td>1.69</td>
<td>1.07</td>
<td>-0.35</td>
<td>-1.98</td>
<td>.51</td>
</tr>
<tr>
<td>20</td>
<td>0.88</td>
<td>1.54</td>
<td>-2.98</td>
<td>-0.41</td>
<td>.55</td>
</tr>
</tbody>
</table>
Table 3

*Item Parameters for the Noncompensatory Model with Rho = .25*

<table>
<thead>
<tr>
<th>Item #</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>E(p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1.38</td>
<td>1.74</td>
<td>-0.79</td>
<td>-0.14</td>
<td>.39</td>
</tr>
<tr>
<td>02</td>
<td>2.40</td>
<td>1.14</td>
<td>0.35</td>
<td>-1.88</td>
<td>.34</td>
</tr>
<tr>
<td>03</td>
<td>1.36</td>
<td>1.50</td>
<td>-1.27</td>
<td>-0.91</td>
<td>.58</td>
</tr>
<tr>
<td>04</td>
<td>1.44</td>
<td>1.45</td>
<td>-0.56</td>
<td>-0.51</td>
<td>.42</td>
</tr>
<tr>
<td>05</td>
<td>1.17</td>
<td>1.94</td>
<td>-2.30</td>
<td>-0.60</td>
<td>.61</td>
</tr>
<tr>
<td>06</td>
<td>1.40</td>
<td>1.66</td>
<td>-1.28</td>
<td>-0.73</td>
<td>.56</td>
</tr>
<tr>
<td>07</td>
<td>1.45</td>
<td>1.40</td>
<td>-1.34</td>
<td>-1.47</td>
<td>.72</td>
</tr>
<tr>
<td>08</td>
<td>1.05</td>
<td>2.47</td>
<td>-3.30</td>
<td>-0.22</td>
<td>.55</td>
</tr>
<tr>
<td>09</td>
<td>1.02</td>
<td>1.09</td>
<td>-2.49</td>
<td>-2.17</td>
<td>.79</td>
</tr>
<tr>
<td>10</td>
<td>1.17</td>
<td>1.47</td>
<td>-1.72</td>
<td>-0.85</td>
<td>.60</td>
</tr>
<tr>
<td>11</td>
<td>1.34</td>
<td>1.34</td>
<td>-0.71</td>
<td>-0.68</td>
<td>.46</td>
</tr>
<tr>
<td>12</td>
<td>1.21</td>
<td>2.06</td>
<td>-1.82</td>
<td>-0.20</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>1.90</td>
<td>0.98</td>
<td>-0.36</td>
<td>-2.80</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>0.93</td>
<td>1.72</td>
<td>-4.65</td>
<td>-1.15</td>
<td>.78</td>
</tr>
<tr>
<td>15</td>
<td>1.29</td>
<td>1.42</td>
<td>-1.08</td>
<td>-0.69</td>
<td>.51</td>
</tr>
<tr>
<td>16</td>
<td>1.84</td>
<td>1.33</td>
<td>-0.27</td>
<td>-1.16</td>
<td>.45</td>
</tr>
<tr>
<td>17</td>
<td>1.97</td>
<td>1.20</td>
<td>-0.66</td>
<td>-2.19</td>
<td>.62</td>
</tr>
<tr>
<td>18</td>
<td>1.15</td>
<td>1.16</td>
<td>-0.28</td>
<td>-0.27</td>
<td>.33</td>
</tr>
<tr>
<td>19</td>
<td>1.80</td>
<td>1.18</td>
<td>-0.35</td>
<td>-1.71</td>
<td>.51</td>
</tr>
<tr>
<td>20</td>
<td>0.98</td>
<td>1.61</td>
<td>-2.57</td>
<td>-0.40</td>
<td>.55</td>
</tr>
</tbody>
</table>
Table 4

*Item Parameters for the Noncompensatory Model with Rho = .50*

<table>
<thead>
<tr>
<th>Item #</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$E(p$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1.52</td>
<td>1.82</td>
<td>-0.66</td>
<td>-0.12</td>
<td>.39</td>
</tr>
<tr>
<td>02</td>
<td>2.48</td>
<td>1.27</td>
<td>0.32</td>
<td>-1.51</td>
<td>.35</td>
</tr>
<tr>
<td>03</td>
<td>1.49</td>
<td>1.63</td>
<td>-1.14</td>
<td>-0.85</td>
<td>.58</td>
</tr>
<tr>
<td>04</td>
<td>1.54</td>
<td>1.54</td>
<td>-0.50</td>
<td>-0.45</td>
<td>.42</td>
</tr>
<tr>
<td>05</td>
<td>1.32</td>
<td>2.04</td>
<td>-1.97</td>
<td>-0.59</td>
<td>.61</td>
</tr>
<tr>
<td>06</td>
<td>1.55</td>
<td>1.79</td>
<td>-1.13</td>
<td>-0.68</td>
<td>.56</td>
</tr>
<tr>
<td>07</td>
<td>1.58</td>
<td>1.55</td>
<td>-1.23</td>
<td>-1.33</td>
<td>.67</td>
</tr>
<tr>
<td>08</td>
<td>1.20</td>
<td>2.51</td>
<td>-2.78</td>
<td>-0.22</td>
<td>.55</td>
</tr>
<tr>
<td>09</td>
<td>1.10</td>
<td>1.17</td>
<td>-2.30</td>
<td>-2.03</td>
<td>.78</td>
</tr>
<tr>
<td>10</td>
<td>1.28</td>
<td>1.56</td>
<td>-1.53</td>
<td>-0.80</td>
<td>.60</td>
</tr>
<tr>
<td>11</td>
<td>1.44</td>
<td>1.43</td>
<td>-0.64</td>
<td>-0.61</td>
<td>.46</td>
</tr>
<tr>
<td>12</td>
<td>1.36</td>
<td>2.12</td>
<td>-1.54</td>
<td>-0.19</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>1.96</td>
<td>1.09</td>
<td>-0.36</td>
<td>-2.39</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>1.03</td>
<td>1.77</td>
<td>-4.07</td>
<td>-1.13</td>
<td>.77</td>
</tr>
<tr>
<td>15</td>
<td>1.39</td>
<td>1.51</td>
<td>-0.97</td>
<td>-0.64</td>
<td>.51</td>
</tr>
<tr>
<td>16</td>
<td>1.95</td>
<td>1.47</td>
<td>-0.26</td>
<td>-0.99</td>
<td>.46</td>
</tr>
<tr>
<td>17</td>
<td>2.08</td>
<td>1.35</td>
<td>-0.63</td>
<td>-1.89</td>
<td>.62</td>
</tr>
<tr>
<td>18</td>
<td>1.21</td>
<td>1.20</td>
<td>-0.23</td>
<td>-0.23</td>
<td>.33</td>
</tr>
<tr>
<td>19</td>
<td>1.89</td>
<td>1.30</td>
<td>-0.34</td>
<td>-1.46</td>
<td>.51</td>
</tr>
<tr>
<td>20</td>
<td>1.08</td>
<td>1.66</td>
<td>-2.23</td>
<td>-0.40</td>
<td>.55</td>
</tr>
</tbody>
</table>
Table 5

*Item Parameters for the Noncompensatory Model with Rho = .75*

<table>
<thead>
<tr>
<th>Item #</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$E(p-value)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1.65</td>
<td>1.92</td>
<td>-0.51</td>
<td>-0.10</td>
<td>.40</td>
</tr>
<tr>
<td>02</td>
<td>2.53</td>
<td>1.43</td>
<td>0.31</td>
<td>-1.14</td>
<td>.35</td>
</tr>
<tr>
<td>03</td>
<td>1.60</td>
<td>1.73</td>
<td>-1.01</td>
<td>-0.77</td>
<td>.58</td>
</tr>
<tr>
<td>04</td>
<td>1.63</td>
<td>1.64</td>
<td>-0.42</td>
<td>-0.39</td>
<td>.43</td>
</tr>
<tr>
<td>05</td>
<td>1.48</td>
<td>2.14</td>
<td>-1.67</td>
<td>-0.57</td>
<td>.61</td>
</tr>
<tr>
<td>06</td>
<td>1.69</td>
<td>1.92</td>
<td>-0.98</td>
<td>-0.62</td>
<td>.56</td>
</tr>
<tr>
<td>07</td>
<td>1.69</td>
<td>1.66</td>
<td>-1.13</td>
<td>-1.21</td>
<td>.66</td>
</tr>
<tr>
<td>08</td>
<td>1.36</td>
<td>2.57</td>
<td>-2.25</td>
<td>-0.22</td>
<td>.55</td>
</tr>
<tr>
<td>09</td>
<td>1.15</td>
<td>1.21</td>
<td>-2.17</td>
<td>-1.93</td>
<td>.78</td>
</tr>
<tr>
<td>10</td>
<td>1.38</td>
<td>1.63</td>
<td>-1.36</td>
<td>-0.76</td>
<td>.60</td>
</tr>
<tr>
<td>11</td>
<td>1.50</td>
<td>1.51</td>
<td>-0.56</td>
<td>-0.54</td>
<td>.46</td>
</tr>
<tr>
<td>12</td>
<td>1.53</td>
<td>2.23</td>
<td>-1.22</td>
<td>-0.19</td>
<td>.49</td>
</tr>
<tr>
<td>13</td>
<td>1.98</td>
<td>1.26</td>
<td>-0.36</td>
<td>-1.99</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>1.15</td>
<td>1.78</td>
<td>-3.60</td>
<td>-1.11</td>
<td>.77</td>
</tr>
<tr>
<td>15</td>
<td>1.47</td>
<td>1.59</td>
<td>-0.85</td>
<td>-0.59</td>
<td>.51</td>
</tr>
<tr>
<td>16</td>
<td>2.03</td>
<td>1.63</td>
<td>-0.23</td>
<td>-0.81</td>
<td>.46</td>
</tr>
<tr>
<td>17</td>
<td>2.15</td>
<td>1.53</td>
<td>-0.61</td>
<td>-1.60</td>
<td>.62</td>
</tr>
<tr>
<td>18</td>
<td>1.24</td>
<td>1.24</td>
<td>-0.18</td>
<td>-0.18</td>
<td>.34</td>
</tr>
<tr>
<td>19</td>
<td>1.94</td>
<td>1.44</td>
<td>-0.33</td>
<td>-1.23</td>
<td>.51</td>
</tr>
<tr>
<td>20</td>
<td>1.17</td>
<td>1.70</td>
<td>-1.92</td>
<td>-0.40</td>
<td>.55</td>
</tr>
</tbody>
</table>
Table 6

*Compensatory Minus Noncompensatory Density Differences in Number-correct Score*

<table>
<thead>
<tr>
<th>Number-correct score (y)</th>
<th>rho</th>
<th>.00</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>.013</td>
<td>.014</td>
<td>.014</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>.015</td>
<td>.012</td>
<td>.009</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>.012</td>
<td>.007</td>
<td>.003</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>.007</td>
<td>.003</td>
<td>.000</td>
<td>-.002</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>.002</td>
<td>-.001</td>
<td>-.002</td>
<td>-.003</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-.003</td>
<td>-.003</td>
<td>-.004</td>
<td>-.003</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>-.006</td>
<td>-.005</td>
<td>-.004</td>
<td>-.003</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-.009</td>
<td>-.007</td>
<td>-.005</td>
<td>-.003</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-.011</td>
<td>-.007</td>
<td>-.005</td>
<td>-.002</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-.012</td>
<td>-.008</td>
<td>-.005</td>
<td>-.002</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-.012</td>
<td>-.008</td>
<td>-.004</td>
<td>-.001</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-.011</td>
<td>-.007</td>
<td>-.004</td>
<td>-.001</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-.009</td>
<td>-.006</td>
<td>-.003</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-.006</td>
<td>-.004</td>
<td>-.002</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-.003</td>
<td>-.002</td>
<td>-.001</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.001</td>
<td>.000</td>
<td>.001</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.005</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.008</td>
<td>.005</td>
<td>.003</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.009</td>
<td>.006</td>
<td>.004</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.008</td>
<td>.006</td>
<td>.003</td>
<td>-.001</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-.005</td>
<td>.003</td>
<td>.001</td>
<td>-.004</td>
<td></td>
</tr>
</tbody>
</table>
Table 7

*Central Moments of Number-correct Scores*

<table>
<thead>
<tr>
<th>MIRT Models</th>
<th>rho</th>
<th>Mean</th>
<th>Second Central Moment</th>
<th>Third Central Moment</th>
<th>Fourth Central Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.00</td>
<td>10.90</td>
<td>25.79</td>
<td>-16.56</td>
<td>1362.83</td>
</tr>
<tr>
<td>Compensatory</td>
<td>.25</td>
<td>10.88</td>
<td>29.40</td>
<td>-20.44</td>
<td>1680.36</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>10.86</td>
<td>32.64</td>
<td>-24.01</td>
<td>1980.03</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>10.84</td>
<td>35.57</td>
<td>-27.27</td>
<td>2262.98</td>
</tr>
<tr>
<td></td>
<td>.00</td>
<td>10.79</td>
<td>20.67</td>
<td>-9.42</td>
<td>946.49</td>
</tr>
<tr>
<td>Noncompensatory</td>
<td>.25</td>
<td>10.78</td>
<td>25.43</td>
<td>-15.86</td>
<td>1336.75</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>10.78</td>
<td>30.12</td>
<td>-24.30</td>
<td>1760.64</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>10.78</td>
<td>34.70</td>
<td>-32.74</td>
<td>2200.57</td>
</tr>
</tbody>
</table>
Table 8

*Ideal Observer Index*

<table>
<thead>
<tr>
<th>Item #</th>
<th>.00</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>.5479</td>
<td>.5397</td>
<td>.5295</td>
<td>.5179</td>
</tr>
<tr>
<td>02</td>
<td>.5311</td>
<td>.5265</td>
<td>.5205</td>
<td>.5128</td>
</tr>
<tr>
<td>03</td>
<td>.5513</td>
<td>.5418</td>
<td>.5307</td>
<td>.5183</td>
</tr>
<tr>
<td>04</td>
<td>.5461</td>
<td>.5377</td>
<td>.5279</td>
<td>.5171</td>
</tr>
<tr>
<td>05</td>
<td>.5421</td>
<td>.5353</td>
<td>.5265</td>
<td>.5157</td>
</tr>
<tr>
<td>06</td>
<td>.5550</td>
<td>.5451</td>
<td>.5332</td>
<td>.5194</td>
</tr>
<tr>
<td>07</td>
<td>.5511</td>
<td>.5419</td>
<td>.5304</td>
<td>.5175</td>
</tr>
<tr>
<td>08</td>
<td>.5243</td>
<td>.5212</td>
<td>.5165</td>
<td>.5102</td>
</tr>
<tr>
<td>09</td>
<td>.5276</td>
<td>.5227</td>
<td>.5162</td>
<td>.5092</td>
</tr>
<tr>
<td>10</td>
<td>.5430</td>
<td>.5351</td>
<td>.5254</td>
<td>.5149</td>
</tr>
<tr>
<td>11</td>
<td>.5435</td>
<td>.5355</td>
<td>.5260</td>
<td>.5156</td>
</tr>
<tr>
<td>12</td>
<td>.5448</td>
<td>.5375</td>
<td>.5281</td>
<td>.5166</td>
</tr>
<tr>
<td>13</td>
<td>.5391</td>
<td>.5246</td>
<td>.5185</td>
<td>.5112</td>
</tr>
<tr>
<td>14</td>
<td>.5124</td>
<td>.5109</td>
<td>.5082</td>
<td>.5048</td>
</tr>
<tr>
<td>15</td>
<td>.5456</td>
<td>.5370</td>
<td>.5271</td>
<td>.5161</td>
</tr>
<tr>
<td>16</td>
<td>.5497</td>
<td>.5411</td>
<td>.5307</td>
<td>.5182</td>
</tr>
<tr>
<td>17</td>
<td>.5442</td>
<td>.5371</td>
<td>.5276</td>
<td>.5162</td>
</tr>
<tr>
<td>18</td>
<td>.5281</td>
<td>.5232</td>
<td>.5175</td>
<td>.5114</td>
</tr>
<tr>
<td>19</td>
<td>.5425</td>
<td>.5352</td>
<td>.5260</td>
<td>.5156</td>
</tr>
<tr>
<td>20</td>
<td>.5292</td>
<td>.5241</td>
<td>.5179</td>
<td>.5108</td>
</tr>
</tbody>
</table>
Table 9

*Average Bias (parameter estimate - true parameter) and Standard Deviation of Bias in Estimates of the Generalized MIRT Model Parameters*

<table>
<thead>
<tr>
<th>Response Data Model</th>
<th>rho</th>
<th>a₁</th>
<th>a₂</th>
<th>d</th>
<th>b₁</th>
<th>b₂</th>
<th>µ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.00</td>
<td>.044</td>
<td>.024</td>
<td>.069</td>
<td>.013</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.042)</td>
<td>(.073)</td>
<td>(.158)</td>
<td>(.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compensatory</td>
<td>.25</td>
<td>.044</td>
<td>.040</td>
<td>.125</td>
<td>.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.047)</td>
<td>(.042)</td>
<td>(.275)</td>
<td>(.052)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>.078</td>
<td>.069</td>
<td>.255</td>
<td>.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.055)</td>
<td>(.081)</td>
<td>(.238)</td>
<td>(.060)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>.098</td>
<td>.113</td>
<td>.787</td>
<td>.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.128)</td>
<td>(.080)</td>
<td>(1.930)</td>
<td>(.094)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncompensatory</td>
<td>.00</td>
<td>-.008</td>
<td>.009</td>
<td></td>
<td>.130</td>
<td>.230</td>
<td>-.199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.099)</td>
<td>(.115)</td>
<td></td>
<td>(.448)</td>
<td>(.354)</td>
<td>(.163)</td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td>-.006</td>
<td>-.004</td>
<td>.250</td>
<td>.254</td>
<td></td>
<td>-.197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.090)</td>
<td>(.083)</td>
<td></td>
<td>(.622)</td>
<td>(.464)</td>
<td>(.144)</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>.039</td>
<td>-.076</td>
<td>.191</td>
<td>.183</td>
<td></td>
<td>-.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.145)</td>
<td>(.104)</td>
<td></td>
<td>(.888)</td>
<td>(.265)</td>
<td>(.125)</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>-.155</td>
<td>-.059</td>
<td>.071</td>
<td>.250</td>
<td></td>
<td>-.288</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.220)</td>
<td>(.105)</td>
<td></td>
<td>(.439)</td>
<td>(.421)</td>
<td>(.175)</td>
</tr>
</tbody>
</table>

*Note: standard deviations are in parentheses*
Figure Captions

Figure 1. Difference Between Compensatory and Noncompensatory True Scores: Rho = .00

Figure 2. Difference Between Compensatory and Noncompensatory True Scores: Rho = .25

Figure 3. Difference Between Compensatory and Noncompensatory True Scores: Rho = .50

Figure 4. Difference Between Compensatory and Noncompensatory True Scores: Rho = .75

Figure 5. Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors: Rho = .00

Figure 6. Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors: Rho = .25

Figure 7. Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors: Rho = .50

Figure 8. Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors: Rho = .75
Difference Between Compensatory and Noncompensatory True Scores:

Rho = .00
Difference Between Compensatory and Noncompensatory True Scores:

\[ \text{Rho} = 0.25 \]
Difference Between Compensatory and Noncompensatory True Scores:

Rho = .50
Difference Between Compensatory and Noncompensatory True Scores:

\[ \text{Rho} = .75 \]
Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors:
\[ \text{Rho} = 0.00 \]
Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors:
Rho = .25
Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors:

\[ \text{Rho} = .50 \]
Absolute Difference Between Compensatory and Noncompensatory Test Information Vectors:

$\text{Rho} = .75$
American College Testing Program/Recease

Dr. Myron Faddi
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington. DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Ms. Paul Foley
Naval Personnel R&D Center
San Diego. CA 92122-6600

Dr. Alfred R. Fregly
AFOSR/UL, Box 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rae 529W
1401 W. Taylor Street
Chicago. IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst. MA 01002

Dr. Drew Gliner
Educational Testing Service
Proinceton, NJ 08541

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3900 O'Hara Street
Pittsburgh. PA 15260

Dr. Sherrie Gott
AFRL/LOME
Brooks AFB, TX 78235-5601

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Michael Haber
DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Prof. Edward Haertel
School of Education
Stanford University
Stanford. CA 94305

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hill South, Room 352
Amherst. MA 01003

Dr. DeWynn Hamsh	
University of Illinois
51 Gerry Drive
Champaign. IL 61820

Dr. Grant Henning
Sensor Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Herter
Naval Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Thomas M. Hirsh
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rowanlino Road
Princeton, NJ 08543

Dr. Paul Horn
677 G Street, #184
Chula Vista, CA 92010

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFRL/LOME
Brooks AFB, TX 78235-5601

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
463 East Daniel Street
Champaign, IL 61820

Dr. Steven Huntz
3-104 Educ. N.
Univ. of Alberta
Edmonton, Alberta
CANADA T6G 2G3

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
University of South Carolina
Columbia, SC 29208

Dr. Kumar Jogi-dev
University of Illinois
Department of Statistics
143 Elm Hall
725 South Wright Street
Champaign, IL 61820

Dr. Douglas H. Jones
1280 Woodburn Court
Toledo, OH 43604

Dr. Brian Junker
Conway-Melone University
Department of Statistics
Schaelex Park
Pittsburgh, PA 15213

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5400

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1100

Prof. John A. Kraus
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. Tae-keun Kim
Department of Psychology
Middle Tennessee State
University
P.O. Box 222
Murfreesboro, TN 37132

Mr. Soo-Hoon Kim
Computer-based Education
Research Laboratory
University of Illinois
Urbana. IL 61801

Dr. G. Gage Kingbury
Portland Public Schools
Research and Evaluation Department
561 North Dixon Street
P. O. Box 3107
Portland, OR 97204-3107

Dr. William Koch
Box 740, Med. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. Richard J. Koubet
Department of Biomedical
& Human Factors
139 Engineering & Math Bldg.
Wright State University
Dayton, OH 45435

Dr. Leonard Kroeker
Naval Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lebovis
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Arlington, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison. WI 53706

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Mr. Rodney Lim
University of Illinois
Department of Psychology
405 E. Daniel Street
Champaign, IL 61820

Dr. Robert L. Lunn
Campus Box 149
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analyses
4401 Ford Avenue
P.O. Box 1628
Alexandria, VA 22302-0288

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. Richard Luecht
ACT
P. O. Box 168
Iowa City, IA 52243
Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 3-1-5
Educational Testing Service
Princeton, NJ 08541

Dr. Cleavon J. Martin
Office of Chief of Naval Operations
Navy Annex, Room 2032
Washington, DC 20350

Dr. James R. McBride
HuntRIO
4630 Elmhurst Drive
San Diego, CA 92120

Dr. Clarence C. McCormick
HQ, USNEMCOM/MEPT
2500 Green Bay Road
North Chicago, IL 60064

Mr. Christopher McIvor
University of Illinois
Department of Psychology
403 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinney
Educational Testing Service
Princeton, NJ 08541

Mr. Alan Mead
ACT
875 Maryland Ave., N.W.
Washington, DC 20036

Dr. Timothy Miller
ECS
P. O. Box 148
Newark, DE 19716

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Mr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code A2
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPM-20
Washington, DC 20300

Dr. Rama Nandakumar
Educational Studies
Walden Hall, Room 213E
University of Delaware
Newark, DE 19716

Library, NPRDC
Code P2011
San Diego, CA 92152-6800

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 861
Department of Educational Psychology & Technology
University of Southern California
Los Angeles, CA 90009-0631

Dr. James B. Olsen
WISC Tests Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Contract N00014-88-C-0256
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Judith Ormson
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Ortenzi
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Peter J. Pashley
Educational Testing Service
Rosecale Road
Princeton, NJ 08541

Wayne M. Paolino
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
NPS Graduation School
Monterey, CA 93943-5026

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Mr. Drew Sands
Mail Stop 3-1-5
Educational Testing Service
Princeton, NJ 08541

Dr. Mary Schmida
APOS/NL, IL 531
Boiling ABF, D.C. 20332-6448

Dr. Keiko Tatsunaka
Educational Testing Service
Mail Stop 0-2-2
Princeton, NJ 08541

American College Testing Program/Reckase
11/09/90

Dr. Ronald Sherry
Psychological & Quantitative Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schmida
AFOSR, DC 20332-6448

Dr. Daniel Schofield
AFOSR, DC 20332-6448

Dr. Mary Schmida
AFOSR, DC 20332-6448

Dr. Dan Segall
Naval Personnel R&D Center
San Diego, CA 92152

Dr. Robie Shealy
University of South Carolina
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Kazuo Shimemizu
7-9-24 Kugenuma-Kagami
Fujisawa 251
JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5510
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Judy Sprey
ACT
P.O. Box 148
Iowa City, IA 52243

Dr. Martha Steinberg
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoffl
Center for Naval Analysis
4401 Ford Ave.
P.O. Box 1533
Alexandria, VA 22332-0268

Dr. William Stouffer
University of Illinois
Department of Experimental Psychology
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariprasad Swaminathan
Laboratories of Psychometrics and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Bob Symms
Naval Personnel R&D Center
Code A2
San Diego, CA 92152-6800

Dr. John Tatarski
APOS/NL, IL 531
Boiling ABF, D.C. 20332-6448

Dr. Kenneth Tatsuoka
Educational Testing Service
Mail Stop 3-1-5
Princeton, NJ 08541