DOCUMENT RESUME

ED 241 580 TM 840 127

AUTHOR McKinley, Robert L.; Reckase, Mark D.
TITLE An Evaluation of One- and Three-Parameter Logistic Tailored Testing Procedures for Use with Small Item Pools.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO ONR83-1
PUB DATE Aug 83
CONTRACT N00014-82-K0716
NOTE 39p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adaptive Testing; *Computer Assisted Testing; Estimation (Mathematics); *Item Banks; *Latent Trait Theory; Mathematical Models; Maximum Likelihood Statistics; Quantitative Tests; Simulation; *Testing
IDENTIFIERS *Ability Parameters; ACT Assessment; *Item Parameters; One Parameter Model; Three Parameter Model

ABSTRACT

A two-stage study was conducted to compare the ability estimates yielded by tailored testing procedures based on the one-parameter logistic (1PL) and three-parameter logistic (3PL) models. The first stage of the study employed real data, while the second stage employed simulated data. In the first stage, response data for 3,000 examinees were obtained for the 40 item ACT Assessment Mathematics Usage subtest. The first 2,000 cases were used to obtain item parameter estimates for both models. Using these estimates, 1PL and 3PL tailored tests were simulated using the response data for the remaining 1,000 cases. Both tailored testing procedures employed maximum likelihood ability estimation and maximum information item selection procedures. The two sets of ability estimates were then compared. In the second stage, response data for 3,000 cases were simulated using the 3PL item parameter estimates from the first stage as true parameters. True abilities were selected from the standard normal distribution. The first 2,000 cases were used for 1PL and 3PL calibration of the items, and the remaining 1,000 cases were used to simulate 1PL and 3PL tailored tests. The two sets of ability estimates were compared to each other and to the true ability parameters. Results of both stages of the study indicated that the 1PL and 3PL tailored tests yielded highly correlated ability estimates, and there was no apparent advantage in terms of ability estimation to using one of the models over the other. Because the 1PL procedure was less expensive to use, it was the recommended model for this application. (Author)

An Evaluation of One- and Three-Parameter Logistic Tailored Testing Procedures for Use with Small Item Pools

Robert L. McKinley
and
Mark D. Reckase

Research Report ONR83-1
August 1983

The American College Testing Program
Resident Programs Department
Iowa City, Iowa 52243

Prepared under Contract No. N00014-82-K0716
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1 REPORT NUMBER ONR83-1	2 GOV'T ACCESSION NO.	3 RECIPIENT'S CATALOG NUMBER
4 TITLE (and Subtitle) An Evaluation of One- and Three-Parameter Logistic Tailored Testing Procedures for use with Small Item Pools		5 TYPE OF REPORT & PERIOD COVERED Technical Report 6 PERFORMING ORG. REPORT NUMBER
7 AUTHOR(s) Robert L. McKinley and Mark D. Reckase		8 CONTRACT OR GRANT NUMBER(s) N00014-82-K0716
9 PERFORMING ORGANIZATION NAME AND ADDRESS The American College Testing Program P.O. Box 168 Iowa City, IA 52243		10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N Proj.: RR042-04 T.A.: 042-04-01 W.U.: NR150-499
11 CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12 REPORT DATE August 1983 13 NUMBER OF PAGES 24
14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15 SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

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

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Latent trait models
Tailored Testing
Computerized Testing

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

An Evaluation of One- and Three-Parameter
Logistic Tailored Testing Procedures
for use with Small Item Pools

ABSTRACT

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A two-stage study was conducted to compare the ability estimates yielded by tailored testing procedures based on the one-parameter logistic (1PL) and three-parameter logistic (3PL) models. The first stage of the study employed real data, while the second stage employed simulated data. In the first stage, response data for 3000 examinees were obtained for the 40 item ACT Assessment Mathematics Usage subtest. The first 2000 cases were used to obtain item parameter estimates for both models. Using these estimates, 1PL and 3PL tailored tests were simulated using the response data for the remaining 1000 cases. Both tailored testing procedures employed maximum likelihood ability estimation and maximum information item selection procedures. The two sets of ability estimates were then compared. In the second stage, response data for 3000 cases were simulated using the 3PL item parameter estimates from the first stage as true parameters. True abilities were selected from the standard normal distribution. The first 2000 cases were used for 1PL and 3PL calibration of the items, and the remaining 1000 cases were used to simulate 1PL and 3PL tailored tests. The two sets of ability estimates were compared to each other and to the true ability parameters. Results of both stages of the study indicated that the 1PL and 3PL tailored tests yielded highly correlated ability estimates, and there was no apparent advantage in terms of ability estimation to using one of the models over the other. Because the 1PL procedure was less expensive to use, it was the recommended model for this application.

CONTENTS

An Evaluation of One- and Three-Parameter
Logistic Tailored Testing Procedures
for use with Small Item Pools

Tailored testing has shown considerable promise as an alternative to conventional paper-and-pencil testing, but before it can be implemented on a widescale basis, a number of issues must be addressed. Tailored testing procedures involve a number of complex components, and there are often a number of alternatives which may be chosen for each. Although there has been considerable research conducted in this area, it is still unclear which of the many alternative components should be used in any particular application. For instance, one important component of tailored testing is the item response theory(IRT) model upon which the procedure is to be based. There are numerous IRT models, several of which have been proposed for use in tailored testing. The purpose of this study was to compare tailored testing procedures based on two of the most popular IRT models, the one-parameter lostistic (1PL) and three-parameter logistic (3PL) models, to determine whether one of the two models is preferable to the other in a tailored achievement testing setting. The tailored testing procedures based on the 1PL and 3PL models were compared on the basis of the ability estimates which were yielded by the procedures. Before reporting the results of the study, it may be helpful to review previous research comparing tailored testing procedures based on these two models.

Comparisons of 1PL and 3PL Tailored Testing Procedures

Several studies have been conducted to compare the use of the 1PL and 3PL models for tailored testing. One such study, reported by Koch and Reckase (1978), was a direct comparison of 1PL and 3PL tailored testing procedures in an application to vocabulary measurement. Both procedures employed maximum likelihood ability estimation techniques, and in both procedures items were selected to maximize the information function at the current ability estimate. The results of this study indicated that both models could be successfully applied to vocabulary ability measurement. The 3PL procedure had a slightly higher reliability (a cross between test-retest and equivalent forms reliabilities) than the 1PL procedure ($r = .77$ for the 3PL procedure, $r = .61$ for the 1PL procedure). However, the 3PL procedure failed to converge to ability estimates in nearly one third of the cases, while nonconvergence was not a serious problem with the 1PL procedure.

In a second study, reported by Koch and Reckase (1979), 1PL and 3PL tailored testing procedures were applied to a multidimensional achievement test. Results of this study indicated very poor performance for both procedures, primarily due to small sample sizes, poor linking procedures, and poor selection of the stepsize and initial ability estimates for the maximum likelihood estimation procedure.

A study reported by McKinley and Reckase (1980) attempted to correct the problems encountered in the Koch and Reckase studies. Close attention was paid to appropriate item parameter linking and selection of the operating characteristics of the procedures. The results of this study indicated that both models could be quite successfully applied to tailored achievement testing if correctly implemented. Both 1PL and 3PL reliabilities were higher than the reliability of a classroom test over the same material. The 3PL procedure yielded better fit to the data than the 1PL procedure, and it also yielded higher test information than the 1PL procedure. This study concluded that for tailored achievement testing the 3PL model was the model of choice. However, the test used in this study was highly multidimensional. It is unclear how generalizable the results are to less multidimensional achievement test.

Urry (1970, 1977) also concluded that the 3PL model was the model of choice. Through a series of simulation studies Urry found that tailored testing becomes less effective when a model with an insufficient number of parameters is used. He concluded that construct valdity decreases as a function of the degree of degeneracy of the model, and the 1PL model was particularly inappropriate for use with multiple-choice items because it did not portray multiple-choice response data with fidelity (Urry, 1977).

This review of previous research indicates that if careful attention is paid to all components of the tailored testing procedure, both 1PL and 3PL tailored testing can be successful. The 3PL model tends to yield higher reliabilities and test information than the 1PL procedure, but is more prone to complications such as nonconvergence. It is also indicated that the 3PL model yields better fit to multidimensional data. Thus, the results of these studies tend to favor the 3PL model. Of course, these results were obtained using relatively large item pools. It is unclear from these studies what results would be obtained using smaller item pools. The purpose of this study was to compare the 1PL and 3PL models in a tailored achievement testing application for which a relatively small item pool is available.

Method

Models

The two models selected for this study were the one-parameter logistic (1PL) and the three-parameter logistic (3PL) models. The 1PL model is given by

$$P(x_{ij}) = \frac{\exp((\theta_j - b_i)x_{ij})}{1+\exp(\theta_j - b_i)} ,$$

where θ_j is the ability parameter for examinee j, b_i is the difficulty parameter for item i, x_{ij} is the observed score (0 or 1) on item i for examinee j, and $P(x_{ij})$ is the probability of response x_{ij} to item i by examinee j. The 3PL model is given by

$$P(x_{ij}=1) = c_i+(1-c_i)\frac{\exp(Da_i(\theta_j - b_i))}{1+\exp(Da_i(\theta_j - b_i))} ,$$

where c_i is the pseudo-guessing parameter for item i, a_i is the discrimination parameter for item i, where $P_i(\theta_j)$ is the probability of a correct response to item i by examinee j, and the remaining terms are as previously defined.

Estimation Programs

For both the 1PL and the 3PL models parameters were estimated using the LOGIST program (Wingersky, Barton, and Lord, 1982). For the 1PL model the pseudo-guessing parameter was held fixed at 0.0. The discrimination parameter was held fixed at a value computed by the LOGIST program. To check the 1PL estimates obtained from LOGIST, they were compared to parameter estimates obtained for the same data using the MAX program (Wright and Panchapakesan, 1969), which was designed for use with the 1PL model. Since the results obtained from the two programs were almost identical, LOGIST was used throughout the study. The LOGIST program was used for both models in order to avoid problems due to different parameter estimate scales. For both models the scales were based on the ability estimate distributions.

Tailored Testing Procedures

Tailored testing procedures have three main components: an item selection routine, an ability estimation technique, and a stopping rule. In this study both the 1PL and 3PL procedures selected items to maximize the value of the information function (Birnbaum, 1968) at the most recent ability estimate. The information for each item at the examinee's current ability estimate was computed, and the item with the greatest information at that ability estimate was administered, with the provision that the information had to be greater than 0.226 for the 1PL procedure and 0.450 for the 3PL procedure. These values were selected on the basis of several trial runs. They were selected so as to yield approximately equal average test lengths for the two models. For both procedures 20 items was the maximum test length allowed.

Prior to testing initial estimates of ability were assigned to set the starting points in the item pool. The initial ability estimates for this study were set to be 0.221 for the 1PL procedure and 0.420 for the 3PL procedure. These values represent difficulty values near the medians of the item pool difficulty parameter distributions. The first item was then selected to maximize information at the initial ability estimate. The response of the examinee to that item was then simulated in the following manner. For the first part of the study, response data came from a fixed length, non-tailored test comprised of all the items in the pool. These items had been administered in paper and pencil form to all of the examinees used in this study. An examinee's response to an item in the tailored tests was the actual response of the examinee to the item on the paper and pencil test. For the second part of the study, simulated response data were generated for each examinee for each item in the pool. These data were generated according to the 3PL model using the 3PL item parameter estimates obtained for the real response data and examinee abilities selected a random from a standard normal distribution. These responses were used regardless of whether a 1PL or 3PL based tailored test were used.

Once the response by an examinee to an item had been obtained, a new estimate of ability was computed by adding a fixed stepsize to the old ability estimate if the response were correct, and by subtracting a fixed stepsize if the response were incorrect. This fixed stepsize procedure was used until a maximum likelihood ability estimate could be obtained (i.e., when both correct and incorrect responses were obtained). The stepsize used was 0.300 for both procedures. Each new item was selected to maximize the information at the new ability estimate, with the restriction that no item could be used more than once.

Two stopping rules were used for the tailored testing procedures. The tests were terminated when there were no items left in the item pool with information at the current ability estimate greater than the minimum specified above, or when 20 items had been administered.

Design

This study employed a two-stage design--one involving the use of real data, and one involving simulated data. In the first stage of the study, response data were obtained for a large sample on a relatively short paper and pencil test. Part of the large sample was then used to calibrate the items on the test using both the 1PL and 3PL models. Using the resulting item parameter estimates, 1PL and 3PL tailored tests were simulated for the examinees not included in the calibration sample. The responses by the examinees to the items in the tailored tests were the same responses they made to the items when taking the paper and pencil test.

In the second stage of the study, the item parameter estimates obtained from the 3PL calibration of the paper and pencil test were used as true parameters, along with the true abilities selected at random from the standard normal distribution, to generate simulated response data to fit the 3PL model. Data were generated for a large sample for all the items from the paper and pencil test. The procedure used for the real data part of the study was then repeated using these simulated data.

Data

For the real data part of the study, response data for the 40 item Mathematics Usage subtest of the ACT Assessment (The American College Testing Program, 1982) were obtained for 3000 cases from the October, 1982 administration of the ACT Assessment (Form 23B). For the second stage of the study, data were simulated for 40 items and 3000 cases. For both stages, then, rather small item pools were used.

Analyses

The analyses performed in this study consisted primarily of computing and comparing correlations. For both the real and the simulation data, the 1PL and 3PL tailored test ability estimates were compared by computing the correlation between them. For the simulation data the two sets of ability estimates obtained from the tailored tests were also compared to the true abilities used to generate the data. Again, the comparisons were performed using correlations.

Results

Real Data Analyses

Item Pool Calibration The first analysis performed on the real data was the calibration of the items for use as a tailored testing item pool. The calibration of the items, which was based on response data for the first 2000 examinees, was performed three different ways. The first two calibrations were performed for the 1PL model using the LOGIST and MAX programs while the third was performed for the 3PL model using LOGIST. The MAX and LOGIST 1PL item difficulty parameter estimates had a correlation of 0.999, as did the ability estimates obtained from the two program. This comparison was performed in order to determine whether the LOGIST program could be used for both models throughout the study. These findings indicated that it could, thus simplifying the problem of placing the estimates from the two models on the same scale.

The item parameter estimate distributions obtained for the two models using LOGIST are shown in Figure 1. These distributions are summarized by the statistics shown in Table 1. As can be seen, most of the 3PL discrimination parameter estimates were .60 or higher, so most of the items were of fairly high quality. From the 3PL difficulty parameter estimate distribution, however, it can be seen that the items are appropriate only for a limited range of ability, since most of the item difficulty estimates fall in the range from -1.0 to 1.75. Most of guessing parameter estimates are .3 or less, with only two items having guessing parameter estimates greater than .3. From these data it would appear that these items actually form a fairly high quality item pool for tailored testing, except for the limitation on the range of difficulty.

For the 1PL model, the LOGIST program assigned to all items a discrimination value of 0.561. The pseudo-guessing parameter was, of course, 0.0. The 1PL difficulty parameter estimate distribution is somewhat different from the 3PL difficulty distribution although the two sets of estimates had a correlation of .88, with the biggest difference being a shift downward of the bulk of the estimates for the 1PL model. Most of the difficulty parameter estimates fall within the same range as for the 3PL model, but there appears to be a shift toward the negative end of that range. Still, for that range the items form an item pool of fairly high quality.

Figure 1

The 1PL and 3PL Item Parameter Estimate
Frequency Distributions for the Real Data

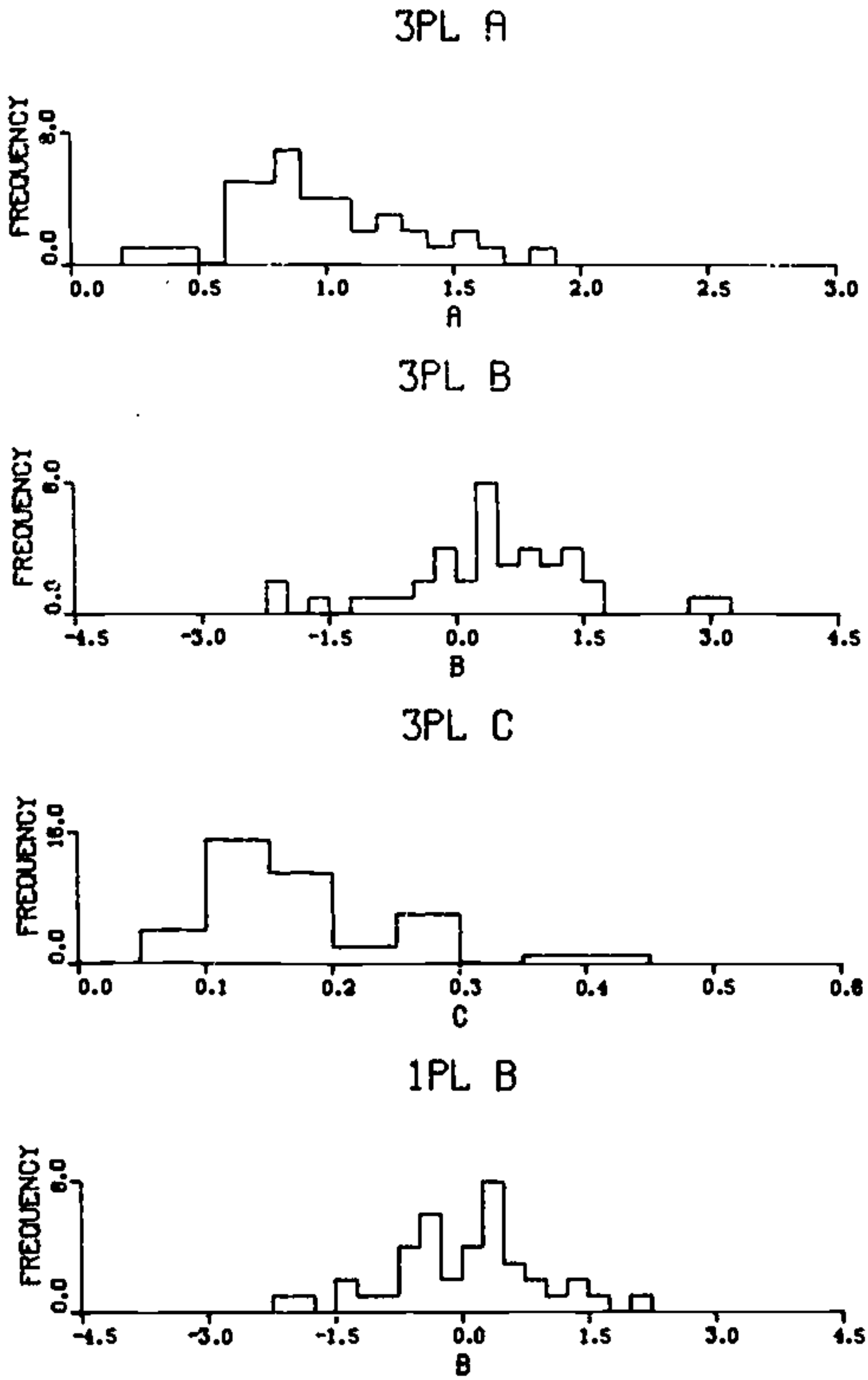


Table 1
Descriptive Statistics of Item Parameter
Estimates for the Real Data

Statistic	1PL	3PL		
	b	a	b	c
Mean	0.03	0.98	0.46	0.17
Median	0.22	0.90	0.41	0.16
S.D.	0.91	0.34	1.10	0.08
Skewness	-0.24	0.40	-0.20	1.14
Kurtosis	0.19	-0.04	0.99	1.19
Low Value	-2.07	0.31	-2.12	0.08
High Value	2.04	1.81	3.15	0.41

Figure 2 shows the test information function for the item pool based on the 1PL item parameter estimates, while Figure 3 shows the test information function based on the 3PL estimates. As can be seen from Figure 3, the 3PL curve is negatively skewed, and is centered around 1.0, thus yielding more information for the positive end of the ability scale. The 1PL curve, on the other hand, is not skewed, and is centered around 0.2. It would appear from this, then, that the 1PL item parameter estimates are appropriate for a wider range for ability than the 3PL estimates are. Of course, the ability scales are not exactly comparable because they are based on different item parameters.

Ability Estimates For those examinees not included in the calibration sample, four different estimates of ability were computed. For each examinee a 1PL and 3PL ability estimate was obtained from simulated tailored test. In addition, ability estimates for each examinee for both models were obtained from LOGIST using the item parameter estimates and the examinee responses from the 40 item paper and pencil test. This made possible not only a comparison of the two tailored testing procedures, but also a comparison of the tailored testing procedures with the paper and pencil tests.

Table 2 summarizes the distributions of the ability estimates obtained for both models from the tailored tests and from the paper and pencil tests. Table 3 shows the intercorrelation matrix for these four sets of ability estimates. As can be seen from these data, the two sets of tailored test ability estimates were similar, with a correlation of 0.77. However, there were some differences in the two distributions.

Figure 2

The Test Information Function for the 1PL Item Parameter Estimates for the Real Data

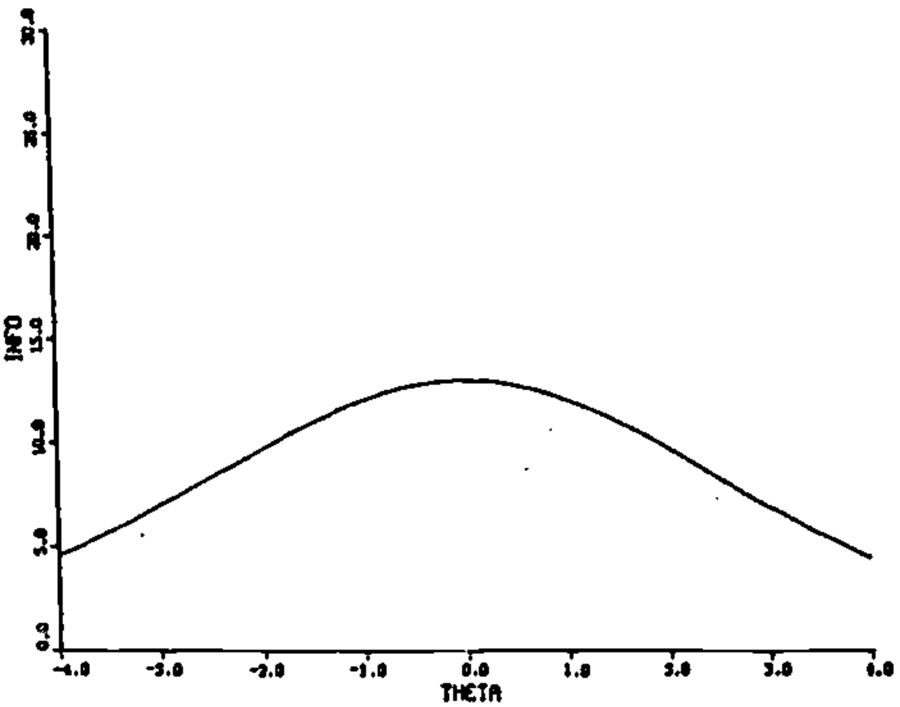
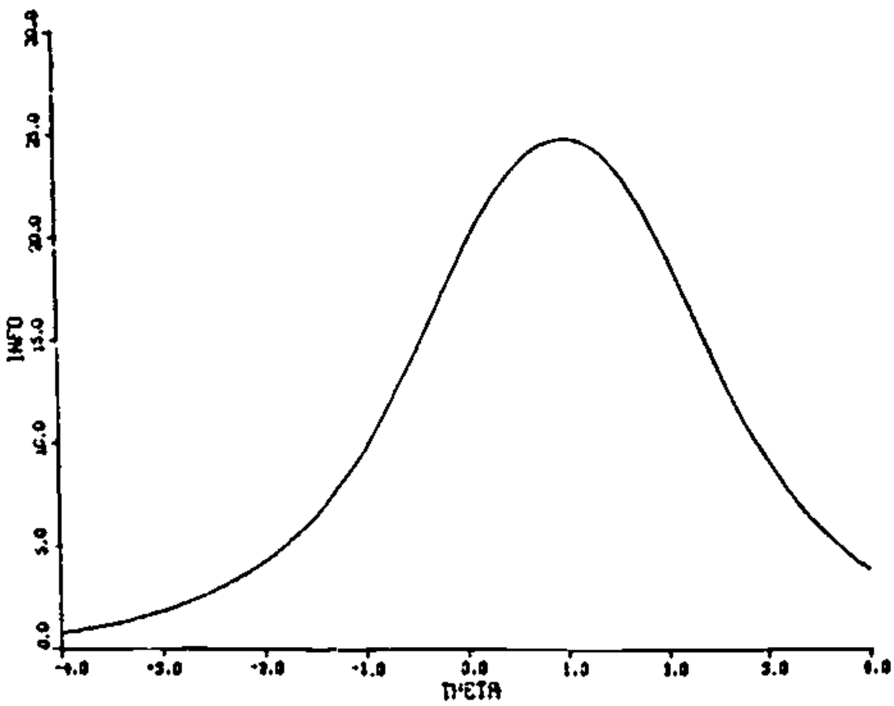


Figure 3

The Test Information Function for the 3PL Item Parameter Estimates for the Real Data



For instance, the skewness value of -0.97 for 3PL ability estimate distribution was significantly different from zero (with a sample size of 1000, the standard error for the skewness coefficient is 0.08), while the 1PL ability estimate distribution was not significantly skewed. Also, the kurtosis value of 1.96 for the 3PL ability estimate distribution was significant (standard error = 0.16), while the kurtosis value of the 1PL ability estimate distribution was not significant.

Table 2
Descriptive Statistics of Ability Parameter Estimates for the Real Data

Statistic	Tailored Tests		Paper and Pencil Tests	
	1PL	3PL	1PL	3PL
Mean	0.15	0.01	0.21	0.11
Median	0.14	0.23	0.16	0.25
S.D.	1.36	1.40	1.13	1.18
Skewness	0.10	-0.97	0.74	-0.35
Kurtosis	0.21	1.96	3.48	4.39
Low Value	-3.65	-4.00	-2.92	-4.00
High Value	6.22	6.42	4.00	4.00
Mean Test Length	12.84	12.16	40.00	40.00
S.D. of Test Length	4.51	4.73	0.00	0.00

Note. For the LOGIST calibrations arbitrary minimums and maximums of -4.00 and 4.00 were set on the ability estimates. The same limits were placed on the tailored tests except in those cases where all items were answered correctly or all were answered incorrectly.

Table 3
Intercorrelation Matrix for Ability Parameter Estimates for the Real Data

Ability Estimate		Tailored Tests 1PL	Tailored Tests 3PL	Paper and Pencil Tests 1PL	Paper and Pencil Tests 3PL
Tailored	1PL	1.00	0.77	0.89	0.87
	3PL		1.00	0.81	0.86
Paper/Pencil	1PL			1.00	0.95
	3PL				1.00

The 1PL and 3PL ability estimates from the paper and pencil test had a correlation of 0.95. Both distributions were leptokurtic (kurtosis = 3.48 for the 1PL estimates, 4.39 for the 3PL estimates), and the two distributions had similar means and standard deviations. The only real difference between these two distributions was that the 3PL distribution was significantly negatively skewed (skewness = -0.35), while the 1PL distribution was significantly positively skewed (skewness = 0.74).

The two sets of tailored test ability estimates were fairly similar to the paper and pencil test ability estimates. The two sets of 1PL estimates had a correlation of 0.89, and the two sets of 3PL estimates had a correlation of 0.86. A comparison of these two correlations via Fisher's r to z transformation yields a $\underline{z} = 2.20$, $\underline{p} < .05$, indicating that the 1PL correlation was significantly higher than the 3PL correlation. Interestingly, the 3PL tailored test ability estimates had a correlation with the 3PL paper and pencil test estimates which was not significantly different from the correlation between the 1PL tailored test ability estimates and the 3PL paper and pencil test ability estimates ($r = 0.86$ for the 3PL estimates, 0.87 for the 1PL estimates). The 1PL tailored test ability estimates did have a significantly higher correlation with the 1PL paper and pencil test estimates than did the 3PL tailored test ability estimates ($r = 0.89$ versus $r = 0.81$).

<u>Average Test Length</u> The average test length for the 1PL tailored tests was 12.8 items, while the average 3PL tailored test was 12.2 items long. This difference is of little or no practical importance, except as an indication that the attempt to produce tests of equal length for the two models was successful. Of some importance is the finding that the 1PL tailored tests required approximately one half of the CPU time required by the 3PL procedures. Of course, if this difference had no signicant impact on response time, then it also is of no practical significance.

<u>Nonconvergence</u> For the 1PL procedure there was no nonconvergence. For the 3PL procedure, however, there was a 4.9% nonconvergence rate. Examinees for whom there was nonconvergence were assigned an ability estimate of 4.0 or -4.0. Of those cases where there was nonconvergence, 96% were at the low end of ability. This is consistent with the finding that the 3PL test information curve was negatively skewed and shifted toward the positive end. Nonconvergence here means that the tailored testing procedure was not able to compute an ability estimate for an examinee. This could

happen because the examinee answered all the items correctly, or all the items incorrectly. It could also happen if the examinee's ability estimate drifted out of the range for which there were appropriate items before both an incorrect and a correct response were obtained. In such a case, the test would be terminated at 20 items, or when both a correct and an incorrect answer were obtained.

Simulation Data Analyses

Item Pool Calibration The first step in the simulation data stage of this study was the generation of data to fit the 3PL model. The true item parameters used for these data were the 3PL item parameter estimates obtained for the real data used in the first part of the study. Data were generated for 3000 cases, using true ability parameters randomly selected from the standard normal distribution. Once these data were generated, the items were calibrated for both the 1PL and 3PL models using the first 2000 cases. The distributions of the obtained item parameter estimates are shown in Figure 4. These distributions are summarized by the statistics shown in Table 4.

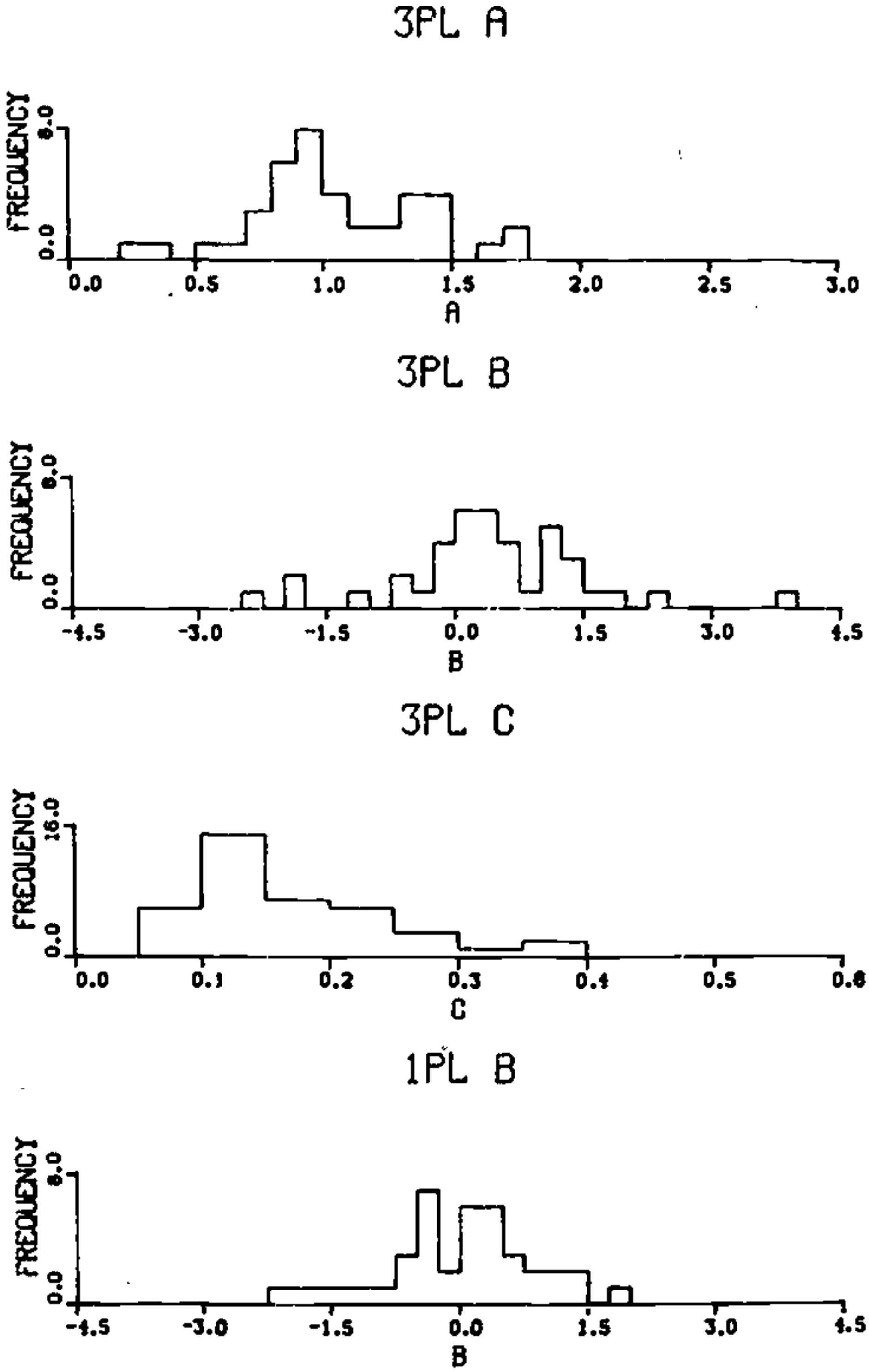
Table 4
Descriptive Statistics of Item Parameter Estimates for the Simulation Data

Statistic	1PL	3PL		
	b	a	b	c
Mean	0.00	1.04	0.41	0.17
Median	0.16	0.96	0.30	0.14
S.D.	0.90	0.34	1.12	0.08
Skewness	-0.31	0.13	0.09	0.94
Kurtosis	0.38	0.11	1.86	1.10
Low Value	-2.20	0.28	-2.27	0.06
High Value	2.00	1.77	3.77	0.40

With few exceptions, these distributions are very much like the distributions of the item parameter estimates obtained for the real data. The only real differences were in the skewness of the 3PL model a-values, which went from slightly positively skewed to not significantly skewed, and the kurtosis of the b-values for the 3PL model, which had an increased kurtosis for the simulation data.

Figure 4

The 1PL and 3PL Item Parameter Estimate
Frequency Distributions for the Simulation Data



One other important difference that was found was that for the 1PL calibration the items were assigned in a-value of 0.60. Since this was higher than the value for the real data (0.56), it was expected that the test information curve for the 1PL model would be higher for the simulation data than for the real data. It was unclear what effect this would have on the simulated 1PL tailored tests, except that it would probably increase the average test length.

Table 5 shows the intercorrelation matrix for the true and estimated item parameters for the simulation data. As can be seen, the 3PL estimates were quite similar to the true parameters. The correlations of the true and estimated 3PL item parameters were 0.89 for the a-values, 0.99 for the b-values, and 0.92 for the c-values. The correlation of the 1PL b-values with the true b-values was 0.88, and the correlation of the 1PL and 3PL b-value estimates was 0.88.

Table 5
Intecorrelation Matrix for the True and Estimated Item Parameters for the Simulation Data

Item		True			1PL Estimates	3PL Estimates		
Parameter		a	b	c	b	a	b	c
True	a	1.00	0.25	0.10	0.45	0.89	0.21	-0.09
	b		1.00	0.40	0.88	0.27	0.99	0.29
	c			1.00	0.11	0.19	0.34	0.92
1PL	b				1.00	0.41	0.88	-0.04
3PL	a					1.00	0.23	0.08
	b						1.00	0.26
	c							1.00

Figures 5 and 6 show the test information curves for the 1PL and 3PL item parameter estimates, respectively. As was the case with the real data, the 3PL information curve is shifted toward the positive end of the ability scale. It is centered around .8. The 1PL curve, on the other hand, is centered around 0.0. The 1PL pool once again appears to be appropriate for a wider range of ability than the 3PL pool is, especially at the lower end of the ability scale. As was predicted from the item calibration results, the 1PL test information curve was higher for the simulation data than for the real data. An unexpected result was that the 3PL test information curve was also higher for the simulation data than for the real data. This was probably

Figure 5

The Test Information Function for the 1PL
Item Parameter Estimates for the Simulation Data

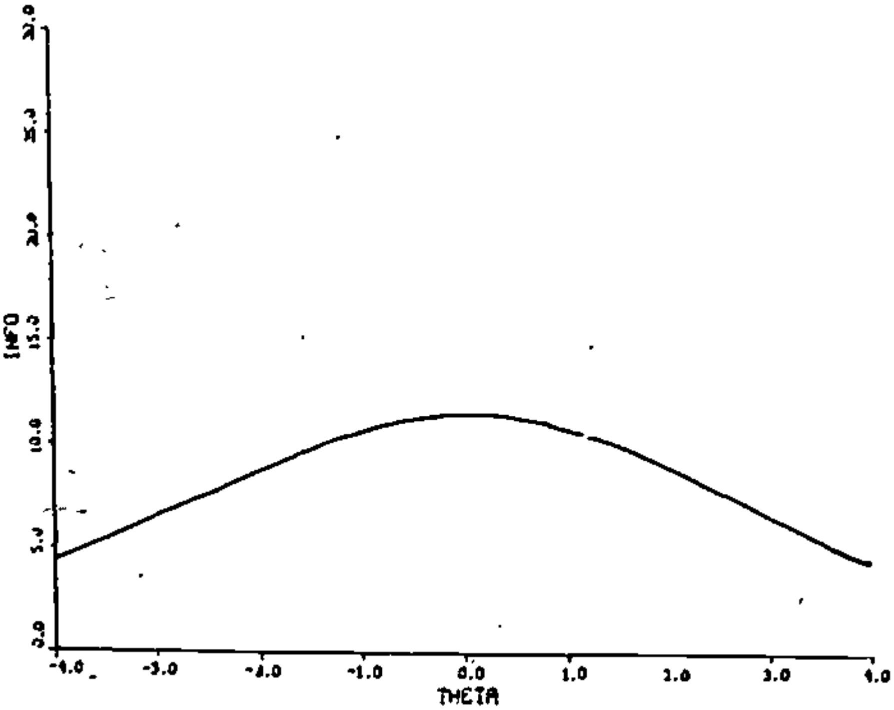
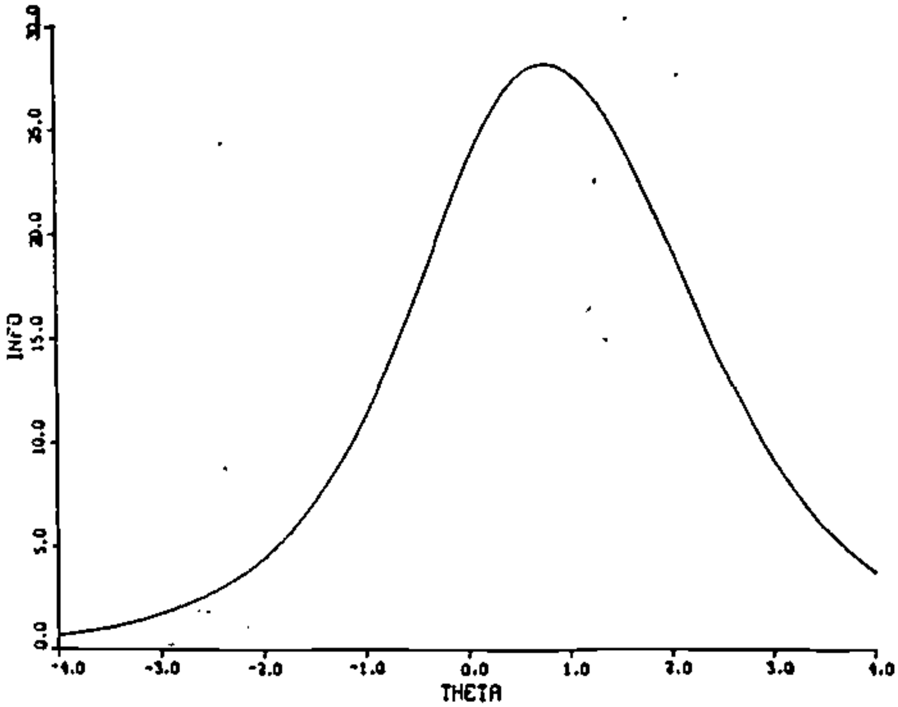


Figure 6

The Test Information Function for the 3PL
Item Parameter Estimates for the Simulation Data



a result of the fact that the simulation data ere generated to fit the 3PL model.

Ability Estimates Four sets of ability estimates were once again computed for the 1000 examinees not included in the calibration sample. For each simulated examinee 1PL and 3PL ability estimates were obtained from the simulated tailored tests as well as from LOGIST runs on the simulated 40 item fixed length test using the item parameter estimates from the calibration of the simulation data. Thus, all the comparisons made with the real data results could be made with the simulation data results. Because these were simulation data and the true ability parameters were known, the ability estimates obtained for these data could also be compared to the true abilities.

The statistics shown in Table 6 summarize the true ability parameter distribution, as well as all of the ability estimate distributions obtained using the simulation data. Table 7 shows the intercorrelation matrix for the true and estimated abilities for the simulation data. The patterns appearing in these data are much like those found for the real data. For these data the correlations are all higher than for the real data, however, with the exception of the correlation between the 1PL and 3PL (simulated) paper and pencil test ability estimates, which was lower for the simulation data (0.928 versus 0.946 for the real data). The 1PL tailored test ability estimates had a correlation of 0.931 with the 1PL simulated paper and pencil test estimates, which was significantly higher than the correlation of 0.826 obtained between the 3PL tailored test estimates and the 1PL paper and pencil test estimates ($\underline{z}$ = 10.954, $\underline{p} < .01$). The 1PL and 3PL tailored test estimates had correlations of 0.920 and 0.854, respectively, with the 3PL paper and pencil test estimates. The difference between these two correlations is significant ($\underline{z}$ = 7.113, $\underline{p} < .01$), indicating that the 1PL correlation was significantly greater than the 3PL correlation.

The inclusion of the true ability parameters in the analyses of the simulation data resulted in a very interesting finding. While the 1PL and 3PL paper and pencil test estimates had correlations with the true parameters that were not significantly different (0.894 for the 3PL estimates, 0.900 for the 1PL estimates), the correlation of the 1PL tailored test ability estimates with the true abilities was significantly higher than the correlation of the 3PL tailored tests ability estimates with the true abilities (r = .883 for the 1PL estimates, 0.816 for the 3PL estimates; $\underline{z}$ = 5.452, $\underline{p} < .01$). This was rather surprising since the simulation data were generated to fit the 3PL model. Just as surprising was the finding that the 1PL tailored test ability estimates had a correlation with the true abilities that was not significantly less than the correlations between the true abilities and the paper and pencil test estimates, despite the fact that the maximum length of the tailored tests was only half the length of the paper and pencil tests.

Table 6
Descriptive Statistics of True and Estimated Abilities for the Simulation Data

Statistic	True	Tailored Tests		Paper and Pencil Tests	
		1PL	3PL	1PL	3PL
Mean	-0.01	-0.08	-0.25	0.02	-0.09
Median	0.00	-0.07	0.00	-0.10	0.03
S.D.	1.04	1.30	1.48	1.11	1.22
Skewness	-0.01	0.32	-0.58	1.11	-0.24
Kurtosis	0.14	0.86	1.52	4.27	4.04
Low Value	-3.82	-3.61	-5.58	-2.47	-4.00
High Value	3.74	6.22	6.42	4.00	4.00
Mean Test Length		17.90	13.51	40.00	40.00
S.D. of Test Length		4.05	5.77	0.00	0.00

Note. For the LOGIST calibrations arbitrary minimums and maximums of -4.00 and 4.00, respectively, were set on the ability estimates. The same limits were placed on the tailored tests except in those cases where all items were answered correctly or all were answered incorrectly.

Table 7
Intercorrelation Matrix for True and Estimated Abilities for the Simulation Data

Ability Estimate		True	Tailored Tests		Paper and Pencil Tests	
			1PL	3PL	1PL	3PL
True		1.00	0.88	0.82	0.90	0.89
Tailored	1PL		1.00	0.81	0.93	0.92
	3PL			1.00	0.83	0.85
P&P	1PL				1.00	0.93
	3PL					1.00

Average Test Length The average test length of the 3PL tailored tests for the simulation data was 13.5 items. The average 1PL tailored test was 17.9 items long. Both of these averages were greater for the simulation data than for the real data as was predicted from the results of the test information curve analyses. The average 3PL test increased by 1.3 while the average 1PL test increased by 5.1. The increased length of the 1PL tests for the simulation data could at least partially explain why the 1PL tailored test estimates had higher correlations with the true abilities and the paper and pencil test estimates than the 3PL tailored test estimates did. Despite the longer average length of the 1PL tailored test, it should be pointed out that the 3PL procedure required half again as much CPU time as the 1PL procedure.

Nonconvergence The 1PL procedure had a .3% nonconvergence rate, while the 3PL procedure had a 5.9% nonconvergence rate. For the 1PL procedure all of the nonconvergence cases (three of them) were at the positive end of the ability scale. For the 3PL procedure 90% of the nonconvergence cases were at the low end of the ability scale. As was the case with the real data, examinees for whom there was nonconvergence were assigned an ability estimate of 4.0 or -4.0.

Discussion

In recent years a number of studies reported in the literature have addressed the issue of whether the 1PL model or the 3PL model should be used in various tailored testing applications. In a tailored achievement testing application, the application of interest here, the research has tended to favor the 3PL model. Because of the inconclusiveness of these studies for applications involving small item pools, and because the 3PL model tends to be more expensive to use, this study was conducted to determine, for a specific application, whether there is sufficient advantage to using the 3PL model to warrant the extra expense. The results of this study will now be discussed, and afterwards some conclusions regarding which model should be used for this application will be presented. First, however, a discussion of the specific application which is of interest in this study will be presented.

The Application

The specific application of interest here has several characteristics which require special consideration. The type of application of concern is an achievement testing application. Achievement testing must be considered in a different light than ability testing because it is learning rather than ability that is being measured. While ability tests generally have learning components, they are constructed to measure a single trait, and as such are usually reasonably unidimensional. Achievement tests, on the other hand, are not specifically directed at a single trait. Moreover, achievement tests often are designed to measure learning in a number of content areas. Therefore, achievement tests typically are not unidimensional, and are often highly multidimensional. The multidimensionality of achievement tests causes problems for IRT, since most IRT models assume unidimensionality.

One way to deal with the dimensionality problem when measuring achievement via IRT is treat the different content areas separately. Individual content areas typically are not unidimensional, but they at least afford a closer approximation to unidimensionalty than do multi-content area tests. Treating content areas separately presents a new problem for tailored testing. A single content area of a test may not include very many items. Tailored testing procedures work best when the item pool has a relatively large number of items, with difficulties spread uniformly over the ability range (Urry, 1977). Building an item pool to meet those specifications, but using only items from a single contant area might be difficult, and certainly would be time-consuming. It seems likely, then, that at least in the early stages of a tailored achievement testing program that treats content areas separately the item pools will be small.

There are at least two other ways to deal with the multidimensionality of achievement tests in a tailored testing application, but at this point neither way is practicable. One way would be to sort the test items into unidimensional subsets, and treat these subsets separately. However, thus far there are no satisfactory procedures for sorting items into unidimensional subsets when the items are dichotomously scored, which achievement test items typically are (Reckase, 1981). Even if sorting could be done, the problem of insufficient items in the pool would still be present.

The other way of dealing with the multidimensionality problem is by using a multidimensional model. Unfortunately, no one has yet developed tailored testing procedures for a multidimensional model. Therefore, this study took the approach of using a unidimensional model with individual content areas. The content area used was the math subtest of the ACT Assessment Program. Using these items, a pool of 40 items was constructed. Using this 40 item pool, a comparison of the 1PL and 3PL models was conducted. The results of that comparison will now be discussed, beginning with the real data part of the study.

Real Data Analyses

Item Pool Calibration Probably the most significant result from the item calibrations was the finding that the 3PL item parameter estimates yielded a test information curve that was negatively skewed and centered around a point on the positive end of the ability scale, while the 1PL item parameter estimates yielded a test information curve that was symmetric and centered around zero. From these results it would be expected that the 3PL tailored tests would tend to terminate prior to convergence for examinees with ability on the lower end of the scale. Such a tendency would not be expected for the 1PL tailored tests.

Ability Estimates The most important finding from the analyses performed on the ability estimates obtained for the real data was that the 1PL model performed as well as the 3PL model without requiring any additional items. The correlation between the 1PL and 3PL tailored testing ability estimates was fairly high (0.772), and the 1PL tailored test estimates were just as highly correlated with the paper and pencil test estimates as were the 3PL tailored test estimates. From these data it appears that there is no advantage to be gained from using the more complex (and expensive) 3PL model.

Average Test Length For the real data tailored test simulations, the average test length for the 1PL and 3PL tests were about the same. This is as it should be, since the information cutoff values for the two procedures were selected to produce tests of equal length.

Nonconvergence There were no cases of nonconvergence for the 1PL tailored test procedure. For the 3PL procedure there was a 4.9% nonconvergence rate. Of those cases where there was nonconvergence, 96% involved examinees at the low end of the ability range. This is consistent with the finding that the 3PL test information curve for the item pool was negatively skewed. Clearly nonconvergence is more of a problem in this case for the 3PL procedure than for the 1PL procedure.

Simulation Data Analyses

Item Pool Calibration What turned out to be one of the most important results of the item calibrations was that for the 1PL calibration LOGIST assigned to the items a common a-value which was higher than that assigned to the items using the real data. This resulted in higher test information for the 1PL model across the ability range. As a result of this, the information cutoff for the 1PL procedure was inappropriately low, which resulted in the tests being longer than expected. The test information curve for the 3PL model was also somewhat higher than for the real data, except at the extremes. This would also be expected to increase the average test length of the 3PL tests, but not as much as for the 1PL tests. The 3PL curve was negatively skewed, as was the case with the real data, which should have once again resulted in some nonconvergence cases at the low end of the ability scale.

Average Test Length As was expected, the average test length increased for both procedures. The 3PL average test length increased by a little over one item, while the average test length for the 1PL procedure increased by about five items. There is no reason to assume that the quality of the 1PL ability estimates would have dramatically decreased had the 1PL tests been shortened by several items, although it would probably have been lower.

Nonconvergence For the simulation data the 3PL nonconvergence rate increased to 5.9%, while the 1PL procedure had a .3% nonconvergence rate. Once again, nonconvergence is clearly a more serious problem for the 3PL procedure than for the 1PL procedure. As was the case for the real data, the bulk of the nonconvergence cases for the 3PL procedure (90%) were at the low end of ability. This is consistent with the results of the test information curve analyses for the simulation data item pools.

Summary and Conclusions

A study was conducted to compare the 1PL and 3PL models in tailored achievement testing application. Both real and simulation data were employed. For the real data, the 1PL procedure was found to yield ability estimates that correlated with paper and pencil test estimates as highly as did the 3PL tailored test ability estimates. The 1PL tests were of about the same average length as were the 3PL tests. For the simulation data, an inappropriately low information cutoff was used for the 1PL procedure, and as a result of the 1PL tests were on the average four to five items longer than the 3PL tests. The 1PL ability estimates were found to be significantly more highly correlated with paper and pencil test estimates than were the 3PL estimates. It was unclear what the results would have been had the 1PL tests been terminated earlier.

The 1PL model is a more appealing model than the 3PL model, since it is simpler to work with, requires smaller sample sizes, and is overall much less expensive to use than the 3PL model. The results of this study indicate that for this type of high quality, small item pool, there is no justification for the added expense and complexity of the 3PL model. For this application, the 1PL model was found to be the model of choice.

REFERENCES

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

Koch, W. R. and Reckase, M. D. A live tailored testing comparison study of the one- and three-parameter logistic models (Research Report 78-1). Columbia, MO: University of Missouri, Department of Educational Psychology, June 1978.

Koch, W. R. and Reckase, M. D. Problems in application of latent trait models to tailored testing (Research Report 79-1). Columbia, MO: University of Missouri, Department of Educational Psychology, September 1979.

McKinley, R. L. and Reckase, M. D. A successful application of latent trait theory to tailored achievement testing (Research Report 80-1). Columbia, MO: University of Missouri, Department of Educational Psychology, February 1980.

Reckase, M. D. The formation of homogeneous item sets when guessing is a factor in item responses (Research Report 81-5). Columbia, MO: University of Missouri, Department of Educational Pyschology, August 1981.

The American College Testing Program. The ACT Assessment, Form 23B. Iowa City, IA: The American College Testing Program, 1982.

Urry, V. W. A Monte Carlo investigation of logist mental test models. (Doctoral dissertation, Purdue University, 1970). Dissertation Abstracts International, 1971, 31, 6319B. (University Microfilms No. 71-9475)

Urry, V. W. Tailored testing: A successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-196.

Wingersky, M. S., Barton, M. A., and Lord, F. M. LOGIST User's Guide. Princeton, NJ: Educational Testing Services, February 1982.

Wright, B. D. and Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Arthur Bachrach
Environmental Stress Program Center
Naval Medical Research Institute
Bethesda, MD 20014

1 Dr. Meryl S. Baker
Navy Personnel R&D Center
San Diego, CA 92152

1 Liaison Scientist
Office of Naval Research
Branch Office, London
Box 39
FPO New York, NY 09510

1 Dr. Alexander Bory
Applied Psychology
Measurement Division
NAMRL
NAS Pensacola, FL 32508

1 Dr. Robert Breaux
NAVTRAEQUIPCEN
Code N-095R
Orlando, FL 32813

1 Dr. Robert Carroll
NAVOP 115
Washington , DC 20370

1 Chief of Naval Education and Training
Liaison Office
Air Force Human Resource Laboratory
Operations Training Division
WILLIAMS AFB, AZ 85224

1 Dr. Stanley Collyer
Office of Naval Technology
800 N. Quincy Street
Arlington, VA 22217

1 CDR Mike Curran
Office of Naval Research
800 N. Quincy St.
Code 270
Arlington, VA 22217

1 Dr. Doug Davis
CNET
Pensacola, FL

Navy

1 Dr. Tom Duffy
Navy Personnel R&D Center
San Diego, CA 92152

1 Mike Durmeyer
Instructional Program Development
Building 90
NET-PDCD
Great Lakes NTC, IL 60088

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
Code P13
NPRDC
San Diego, CA 92152

1 Dr. Cathy Fernandes
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jim Hollan
Code 14
Navy Personnel R & D Center
San Diego, CA 92152

1 Dr. Ed Hutchins
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William L. Maloy (02)
Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr William Montague
NPRDC Code 13
San Diego, CA 92152

Navy

1 Bill Nordbrock
1032 Fairlawn Ave.
Libertyville, IL 60048

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

1 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

6 Personnel & Training Research Group
Code 442PT
Office of Naval Research
Arlington, VA 22217

1 Special Asst. for Education and Training (OP-01E)
Rm. 2705 Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USN (Ph.D)
CNET (N-432)
NAS
Pensacola, FL 32508

1 Dr. Bernard Rimland (01C)
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

1 Dr. Robert G. Smith
Office of Chief of Naval Operations
OP-987H
Washington, DC 20350

1 Dr. Alfred F. Smode, Director
Training Analysis & Evaluation Group
Dept. of the Navy
Orlando, FL 32813

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frederick Steinheiser
CNO - OP115
Navy Annex
Arlington, VA 20370

Navy

1 Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Edward Wegman
Office of Naval Research (Code 411SSP)
800 North Quincy Street
Arlington, VA 22217

1 Dr. Ronald Weitzman
Naval Postgraduate School
Department of Administrative Sciences
Monterey, CA 93940

1 Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152

Marine Corps

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 Director, Office of Manpower Utilization
HQ, Marine Corps (MPU)
BCB, Bldg. 2009
Quantico, VA 22134

1 Headquarters, U. S. Marine Corps
Code MPI-20
Washington, DC 20380

1 Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

1 Major Frank Yohannan, USMC
Headquarters, Marine Corps
(Code MPI-20)
Washington, DC 20380

Army

1 Technical Director
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Mr. James Baker
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milton S. Katz
Training Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Harold F. O'Neil, Jr.
Director, Training Research Lab
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Commander, U.S. Army Research Institute
for the Behavioral & Social Sciences
ATTN: PERI-BR (Dr. Judith Orasanu)
5001 Eisenhower Avenue
Alexandria, VA 20333

1 Joseph Psotka, Ph.D.
ATTN: PERI-1C
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Army

1 Mr. Robert Ross
U.S. Army Research Institute for the
Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joyce Shields
Army Research Institute for the
Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

1 Dr. Robert Wisher
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

1 Technical Documents Center
Air Force Human Resources Laboratory
WPAFB, OH 45433

1 U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

1 Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Dr. T. M. Longridge
AFHRL/OTE
Williams AFB, AZ 85224

1 Mr. Randolph Park
AFHRL/MOAN
Brooks AFB, TX 78235

1 Dr. Roger Pennell
Air Force Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Air Force

1 3700 TCHTW/TTGHR
2Lt Tallarigo
Sheppard AFB, TX 76311

1 Lt. Col James E. Watson
HQ USAF/MPXOA
The Pentagon
Washington, DC 20330

1 Major John Welsh
AFMPC
Randolph AFB, TX

1 Dr. Joseph Yasatuke
AFHRL/LRT
Lowry AFB, CO 80230

Department of Defense

12 Defense Technical Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC

1 Dr. Craig I. Fields
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

1 Jerry Lehnus
HQ MEPCOM
Attn: MEPCT-P
Fort Sheridan, IL 60037

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Dr. Wayne Sellman
Office of the Assistant Secretary
of Defense (MRA & L)
2B269 The Pentagon
Washington, DC 20301

1 Major Jack Thorpe
DARPA
1400 Wilson Blvd.
Arlington, VA 22209

Civilian Agencies

1 Dr. Susan Chipman
Learning and Development
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E Street NW
Washington, DC 20415

1 Mr. Thomas A. Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Private Sector

1 Dr. James Algina
University of Florida
Gainesville, FL 326

1 Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

1 1 Psychological Research Unit
NBH-3-44 Attn: Librarian
Northbourne House
Turner ACT 2601
AUSTRALIA

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

1 Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
Israel

1 Dr. R. Darrell Bock
Department of Education
University of Chicago
Chicago, IL 60637

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52243

1 Dr. Ernest R. Cadotte
307 Stokely
University of Tennessee
Knoxville, TN 37916

1 Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Private Sector

1 Dr. Hans Crombag
Education Research Center
University of Leyden
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

1 Dr. Dattprasad Divgi
Syracuse University
Department of Psychology
Syracuse, NE 33210

1 Dr. Fritz Drasgow
Department of Psychology
University of Illinois
603 E. Daniel St.
Champaign, IL 61820

1 Dr. Susan Embertson
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
Lawrence, KS 66045

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Benjamin A. Fairbank, Jr.
McFann-Gray & Associates, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

1 Dr. Leonard Feldt
Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Univ. Prof. Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

1 Professor Donald Fitzgerald
University of New England
Armidale, New South Wales 2351
AUSTRALIA

Private Sector

1 Dr. Dexter Fletcher
WICAT Research Institute
1875 S. State St.
Orem, UT 22333

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

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
3939 O'Hara Street
PITTSBURGH, PA 15260

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

1 Dr. Ron Hambleton
School of Education
University of Massachusetts
Amherst, MA 01002

1 Dr. Delwyn Harnisch
University of Illionis
242b Education
Urbana, IL 61801

1 Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

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

1 Dr. Jack Hunter
2122 Coolidge St.
Lansing, MI 48906

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

Private Sector

1 Dr. Douglas H. Jones
Advanced Statistical Technologies Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

1 Professor John A. Keats
Department of Psychology
The University of Newcastle
N.S.W. 2308
AUSTRALIA

1 Dr. William Koch
University of Texas-Austin
Measurement and Evaluation Center
Austin, TX 78703

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

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

1 Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
Netherlands

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Mr. Phillip Livingston
Systems and Applied Sciences Corporatio
6811 Kenilworth Avenue
Riverdale, MD 20840

1 Dr. Robert Lockman
Center for Naval Analysis
200 North Beauregard St.
Alexandria, VA 22311

Private Sector

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

1 Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

1 Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

1 Dr. Scott Maxwell
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556

1 Dr. Samuel T. Mayo
Loyola University of Chicago
820 North Michigan Avenue
Chicago, IL 60611

1 Mr. Robert McKinley
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

1 Dr. Barbara Means
Human Resources Research Organization
300 North Washington
Alexandria, VA 22314

1 Dr. Robert Mislevy
711 Illinois Street
Geneva, IL 60134

1 Dr. Allen Munro
Behavioral Technology Laboratories
1845 Elena Ave., Fourth Floor
Redondo Beach, CA 90277

1 Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73059

1 Dr. Melvin R. Novick
356 Lindquist Center for Measurment
University of Iowa
Iowa City, IA 52242

Private Sector

1 Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

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

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

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

1 Dr. Thomas Reynolds
University of Texas-Dallas
Marketing Department
P. O. Box 688
Richardson, TX 75080

1 Dr. Lawrence Rudner
403 Elm Avenue
Takoma Park, MD 20012

1 Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Frank L. Schmidt
Department of Psychology
Bldg. GG
George Washington University
Washington, DC 20052

1 Dr. Walter Schneider
Psychology Department
603 E. Daniel
Champaign, IL 61820

Private Sector

1 Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

1 DR. ROBERT J. SEIDEL
INSTRUCTIONAL TECHNOLOGY GROUP
HUMRRO
300 N. WASHINGTON ST.
ALEXANDRIA, VA 22314

1 Dr. Kazuo Shigemasu
University of Tohoku
Department of Educational Psychology
Kawauchi, Sendai 980
JAPAN

1 Dr. Edwin Shirkey
Department of Psychology
University of Central Florida
Orlando, FL 32816

1 Dr. William Sims
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Robert Sternberg
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

1 Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

Private Sector

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Lab
252 Engineering Research Laboratory
Urbana, IL 61801

1 Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. Robert Tsutakawa
Department of Statistics
University of Missouri
Columbia, MO 65201

1 Dr. J. Uhlaner
Uhlaner Consultants
4258 Bonavita Drive
Encino, CA 91436

1 Dr. V. R. R. Uppuluri
Union Carbide Corporation
Nuclear Division
P. O. Box Y
Oak Ridge, TN 37830

1 Dr. David Vale
Assessment Systems Corporation
2233 University Avenue
Suite 310
St. Paul, MN 55114

1 Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08540

1 Dr. Michael T. Waller
Department of Educational Psychology
University of Wisconsin--Milwaukee
Milwaukee, WI 53201

Private Sector

1 Dr. Brian Waters
HumRRO
300 North Washington
Alexandria, VA 22314

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90007

1 Wolfgang Wildgrube
Streitkraefteamt
Box 20 50 03
D-5300 Bonn 2
WEST GERMANY

1 Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

1 Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940